

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
9 December 2010 (09.12.2010)

PCT

(10) International Publication Number
WO 2010/139858 A1

(51) International Patent Classification:

C12N 15/70 (2006.01) C12N 9/00 (2006.01)
C12N 9/02 (2006.01) C07K 14/00 (2006.01)
C12N 15/00 (2006.01) C12N 5/00 (2006.01)

(21) International Application Number:

PCT/FI2010/050448

(22) International Filing Date:

2 June 2010 (02.06.2010)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

20095615 2 June 2009 (02.06.2009) FI

(71) Applicant (for all designated States except US): **OULUN YLIOPISTO** [FI/FI]; Pentti Kaiteran katu 1, FIN-90014 Oulun Yliopisto (FI).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **RUDDOCK, Lloyd** [GB/FI]; Sumukellonkuja 12, FIN-90520 Oulu (FI).

(74) Agent: **BERGGREN OY AB**; P.O. Box 16 (Antinkatu 3 C), FIN-00101 Helsinki (FI).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

— with sequence listing part of description (Rule 5.2(a))

(54) Title: METHOD FOR PRODUCING NATIVELY FOLDED PROTEINS IN A PROKARYOTIC HOST

(57) Abstract: The present invention relates to a method for producing a protein of interest containing one or more disulfide bonds in its native state. The method comprises that a prokaryotic host cell is genetically engineered to express the protein of interest and a sulfhydryl oxidase in the cytoplasm of the host cell. The protein of interest is formed in a soluble form and contains disulfide bonds due to the presence of the sulfhydryl oxidase in the cytoplasm of said host cell. The present invention relates also to a prokaryotic host cell and a vector system for producing a protein of interest containing natively folded disulfide bonds.



WO 2010/139858 A1

Method for producing natively folded proteins in a prokaryotic host

Field of the invention

5 This invention relates to a method, a host cell and a vector system for producing a protein of interest containing one or more disulfide bonds in its native state. In particular, the invention relates to a method, a host cell and a vector system for producing such proteins in a prokaryotic host.

Description of Related Art

10 Many proteins and enzymes of biotechnological importance contain structure stabilizing disulfide bonds, with an estimated one third of all human proteins folding in the endoplasmic reticulum (ER) and acquiring disulfide bonds there. This includes most proteins which get secreted or end up on the outer membrane. Since any two cysteine residues in a protein have the potential to form a disulfide bond, the correct formation of native disulfide bonds is not trivial. Hence, it is unsurprising that native disulfide
15 bond formation is often the rate-limiting step in the folding of proteins *in vitro* and *in vivo*.

The process of native disulfide bond formation in the endoplasmic reticulum (ER), periplasm or inter-membrane space of mitochondria is known to be catalysed by several families of enzymes. However, while some of the participants in the cellular process
20 are known, their precise individual roles are still largely confused.

What is known is that native disulphide bond formation can occur via multiple parallel pathways (see **Figure 1**). The current state of the art says that the major pathway for disulfide bond formation in the ER is via oxidation of protein disulfide isomerase (PDI) family members by the sulfhydryl oxidase activity of the gene product(s) of *ERO1*
25 family members. PDI in turn then introduces disulfide bonds into folding proteins.

Parallel pathways, for example via glutathione, are also possible. Similarly in the inter-membrane space of mitochondria a similar pathway exists via the oxidation of Mia40 by the sulfhydryl oxidase Erv1p. Mia40 in turn introduces disulfide bonds into folding proteins. A similar pathway exists in the periplasm of prokaryotes via DsbA
30 and DsbB except that DsbB is not a sulfhydryl oxidase.

During disulfide bond formation many non-native disulfide bonds may be formed. Hence *in vitro* in a glutathione redox buffer, and probably *in vivo* for many proteins, the rate-limiting step for native disulfide bond formation is late-stage isomerization reactions, where disulfide bond isomerization is linked to conformational changes in
5 protein substrates with substantial regular secondary structure. These steps are catalysed by thiol-disulfide isomerases, in particular in the ER by proteins belonging to the PDI-family and in the periplasm of prokaryotes by DsbC and DsbG.

Currently proteins that contain disulfide bonds are difficult for the biotech industry to produce on a large scale. The most common route is to produce these proteins in the
10 cytoplasm of *E.coli*. Here there are no mechanisms for disulfide bond formation. Due to this the recombinant proteins are unable to attain their native conformation and form insoluble inclusion bodies. Inclusion body refolding is a widely studied and widely patented field. However, it is costly, complex and generally inefficient. Alternative routes for producing disulfide bonded proteins also have drawbacks.

15 I. Disulfide bond formation in the periplasm of *E.coli*. While native *E.coli* disulfide bond containing proteins fold efficiently in the periplasm, the yields of heterologously expressed proteins are often very low, in part due to the small size of the periplasm. In addition, the outer membrane of *E.coli* is freely diffusible to most small molecules which means that the biophysical environment of
20 the periplasm is very dependent on the external media.

II. Disulfide bond formation in the cytoplasm of modified *E.coli*. *E.coli* has two pathways to ensure that its cytoplasm is reducing: i) using thioredoxins / thioredoxin reductases and ii) using glutathione / glutaredoxin / glutathione reductase (see **Figure 2**). When both pathways are knocked out, for example in
25 the commercial origami or rosetta-gami strains (Novagen®), the cytoplasm is less reducing and disulfide bonds form in proteins. However, disulfide bond formation is still slow and inefficient. In addition, these strains are less genetically stable and grow significantly more slowly than wild type strains. While some disulfide bond containing proteins can be formed in the cytoplasm of origami or rosetta-gami (and equivalent), the yields of most proteins are below
30 that required for commercial production.

III. Disulfide bond formation in the ER of eukaryotic organisms such as *S.cerevisiae*, *Pichia pastoris*, insect cell culture or even mammalian cell culture is

much more efficient than that in bacteria. However, there is the corresponding increase in costs associated with the growth of eukaryotic organisms and problems associated with the large scale production of proteins in cell culture.

5 WO 9907727 A1 (or U.S. Patent No. 6,361,964) describes a method of increasing disulfide bond formation in a protein by expressing the protein in a host cell that also expresses an isolated nucleic acid that encodes an Ero1 polypeptide or optionally an Ero1 polypeptide together with a protein disulfide isomerase.. The Ero1 polypeptide is suggested to be for use in eukaryotic expression systems, wherein Ero1 aids folding in the ER, or *in vitro* refolding reactions.

10 WO 2005061718 A1 describes a method for producing a heterologous protein in a host cell utilizing a 2 micron-family plasmid. The host cell is cultured in such conditions that allow the expression of the gene encoding the fungal molecular chaperone or protein folding catalyst, such as Ero1 or protein disulfide isomerase, and the gene encoding a heterologous protein. The method uses 2 micron-family plasmids which
15 can be applied only to eukaryotic cells. Both the protein and the molecular chaperone or protein folding catalyst such as Ero1 are targeted to the ER.

WO 2009058956 A1 describes a method for expressing a protein in a filamentous fungal host, wherein Ero1 can be co-expressed with the desired protein. Ero1 facilitates the folding of the desired protein in the ER.

20 KR 20070041166 A describes a method for producing a protein in the ER of a yeast cell, wherein Ero1 and protein disulfide isomerase are simultaneously expressed in a host cell.

Furthermore, U.S. Patent Application No. 2007/0193977 describes a method for producing a desired protein in a host cell comprising recombinant genes for a first and a
25 second chaperone and for the desired protein. A preferred chaperone is PDI, in particular from fungal or mammalian origin. According to the publication various host expression systems including yeast, bacteria and mammalian cells may be used. Bacterial host cells are mentioned to be useful for cloning purposes.

Examples of commercial proteins that are produced by these various routes include
30 insulin, tissue plasminogen activator, growth hormones and single chain antibodies.

None of the mentioned patent publications disclose a prokaryotic expression system for the production of natively folded disulfide bond containing proteins.

Summary

5 It is an object of the invention to provide a method for producing a natively folded disulfide bond containing protein in a prokaryotic host.

It is also an object of the invention to provide a host cell for producing a natively folded disulfide bond containing protein in a prokaryotic host.

Further, it is also an object of the present invention to provide a vector system for producing a natively folded disulfide bond containing protein in a prokaryotic host.

10 These and other objects together with the advantages thereof over known methods, hosts and vectors, are achieved by the present invention as hereinafter described and claimed.

The invention is based on the use of genetic engineering of prokaryotic host cells to cytoplasmically express a sulfhydryl oxidase.

15 In one aspect, the present invention provides a method for producing a protein of interest containing one or more disulfide bonds in its native state in a prokaryotic host.

In one embodiment the invention provides a method which comprises that a prokaryotic host cell is genetically engineered to express the protein of interest and a sulfhydryl oxidase in the cytoplasm of the host cell, said protein of interest being formed in a
20 soluble form and containing disulfide bonds due to the presence of the sulfhydryl oxidase in the cytoplasm of said host cell.

In one embodiment of the invention the protein product comprising the protein of interest is recovered from the cell culture or from the host cells and optionally purified.

In another aspect, the present invention provides a prokaryotic host cell for producing
25 a protein of interest containing one or more disulfide bonds in its native state in a prokaryotic host.

In one embodiment the invention provides a prokaryotic host cell for producing a protein of interest containing natively folded disulfide bonds, which comprises that the

host cell is genetically engineered to express a sulfhydryl oxidase and a protein of interest in the cytoplasm of the host cell.

In one preferred embodiment of the invention the sulfhydryl oxidase is co-expressed with the protein of interest.

- 5 In another preferred embodiment the sulfhydryl oxidase is expressed prior to the protein of interest.

In one further preferred embodiment, the prokaryotic host cells are engineered to express also a thiol-disulfide isomerase in the cytoplasm.

- 10 In one still further preferred embodiment the thiol-disulfide isomerase is the eukaryotic ER-resident enzyme protein disulfide isomerase (PDI).

In one further preferred embodiment the thiol-disulfide isomerase is DsbC, which is usually targeted to the periplasm of prokaryotes.

- 15 These co-expressions may be used separately or in combination with any of the other variations described herein, so long as a sulfhydryl oxidase is expressed in the host cell.

The host can be any prokaryotic host. In one embodiment of the invention, the prokaryotic host is a bacterial host, in one specific embodiment a gram-negative host, such as *E. coli*.

- 20 In one further preferred embodiment, expression of the disulfide bond containing protein of interest is achieved by expressing the protein as a genetic fusion with a fusion partner. This may be used separately or in combination with any of the other variations described herein, so long as a sulfhydryl oxidase is expressed in the host cell.

- 25 In one further embodiment the host cell may be deficient in thioredoxin reductase or glutathione reductase activity. This may be caused by having a mutation or deletion in *trxB* gene and/or a mutation or deletion in *gor* gene. These deficiencies may be used separately or in combination with any of the other variations described herein, so long as a sulfhydryl oxidase is expressed in the host cell.

In still further embodiments, codon usage for the nucleotide sequence encoding the protein may be optimized for expression in a particular host or the host codon usage

manipulated so as to optimize expression of the recombinantly expressed protein. This may be done for example by supplementing the expression host's tRNA levels by production of tRNA species encoded by engineered plasmids. Codon usage may also be optimized for the sulfhydryl oxidase and/or the thiol-disulfide isomerase.

- 5 These modifications may be used separately or in combination with any of the other variations described herein, so long as a sulfhydryl oxidase is expressed in the host cell.

The present invention provides also a vector system, comprising one or more vectors encoding sulfhydryl oxidase and the protein of interest, and optionally also, a thiol-disulfide isomerase. The proteins may be separately inducible.

10

Hence the invention has possible alternative solutions, all with the common factor of cytoplasmic expression of a sulfhydryl oxidase.

The method of the present invention results in the formation of native disulfide bonds in a significant proportion. The method is particularly suitable for expressing proteins where a biological activity is dependent on the formation of one or more intra- or inter-chain disulfide bonds.

15

In the following, the invention will be examined more closely with the aid of a detailed description and with reference to some working examples.

Brief description of the figures:

- 20 **Figure 1** shows pathways for disulfide bond formation in: Panel A: the endoplasmic reticulum; Panel B: The inter-membrane space of mitochondria (protein names given for the system in *Saccharomyces cerevisiae*); Panel C: The periplasm of prokaryotes (protein names given for the system in *Escherichia coli*).

- 25 **Figure 2** shows the pathways for reducing disulfide bonds in the cytoplasm of *E.coli*. Gene names are given in parentheses. In commercial strains for the production of disulfide bonded proteins, such as origami and rosetta-gami (Novagen ®), both the thioredoxin and glutathione pathways are knocked out by *gor* and *trxB* deletions or knock-outs by transposon mutagenesis.

- 30 **Figure 3** shows protein sequences of representative members of the ERO1 family of sulfhydryl oxidases (EC 1.8.4.-). The cleavable signal sequences are underlined.

Figures 4A and 4B show protein sequences of representative members of the ERV/ALR sulfhydryl oxidase domain containing sulfhydryl oxidases (Enzyme classification: EC 1.8.3.2). The ERV/ALR domain sequences for each, as currently defined in UniProt, are underlined. Figure 4B shows also protein sequences of thiol-disulfide isomerases (EC 5.3.4.1) PDI and DsbC. The signal sequences are underlined.

Figure 5 shows plasmids pET23 and pLysS as modified for producing natively folded disulfide bond containing proteins in these studies. The order of the genes on polycistronic vectors does not affect the ability to co-expression of the proteins.

Figure 6 shows representative SDS-PAGE from expression of the luminal domain of human tissue factor with co-expression with a sulfhydryl oxidase in LB media at 25°C. Lane 2 shows the total *E.coli* lysate, lane 1 shows the soluble fraction. The position of the luminal domain of tissue factor is marked with an arrow.

Figure 7 shows measurements for the activity of *E.coli* alkaline phosphatase (PhoA) expressed in the cytoplasm of *E.coli* in LB media at 30 °C.

Figure 8 shows measurements for the activity of *E.coli* phytase (AppA) expressed in the cytoplasm of *E.coli* strains in LB media at 30 °C.

Figure 9 shows SDS-PAGE from expression of the human CSF3 (lanes 2-7) and BMP4 (lanes 8-13) as fusion proteins with MBP in the cytoplasm of *E.coli* with co-expression with a sulfhydryl oxidase in LB media at 30 °C and co-expression of PDI (lanes 4,5,10,11) or DsbC (lanes 6,7,12,13). Even numbered lanes show the total *E.coli* lysates, odd numbered lanes show the soluble fractions (except lane 1 which shows molecular weight markers). The position of the MBP fusion proteins CSF3 and BMP4 are marked with arrows.

Figure 10 shows analysis of human kringle tPA in various systems in LB media at 30 °C. Panel A) SDS-PAGE from expression of the human kringle tPA as a fusion protein with MBP in the cytoplasm of *E.coli*. BL21 pLysS RARE (BL) or rosetta-gami (RG) or origami (OG), with co-expression with a sulfhydryl oxidase (SO) and DsbC. Alternate lanes show the total *E.coli* lysates (T), and the soluble fractions (S). The position of the MBP-kringle tPA fusion protein is marked with an arrow. Panel B) shows measurements for the activity of human kringle tPA. Panels C-F) show the results from measurements for the activity of the MBP-kringle t-PA fusion protein.

Figure 11 shows analysis of BPTI production in origami + pre-expression of the sulfhydryl oxidase and PDI in LB media at 30°C. Panel A) SDS-PAGE from expression of BPTI wild type (lanes 1 and 2), along with the C49A/C73A (lanes 3 and 4) and C65A/C86A (lanes 5 and 6) mutants, which should contain only 2 disulfide bonds, in the cytoplasm of origami *E.coli* with pre-expression of the sulfhydryl oxidase (SO) and PDI. Odd numbered lanes show the total *E.coli* lysates, even numbered lanes show the soluble fractions. The position of BPTI is marked with an arrow. Panel B) Purification of the C49A/C73A or C65A/C86A mutants of BPTI using immobilized metal affinity chromatography and subsequent analysis by reducing (lanes 1 and 3) and non-reducing (lanes 2 and 4) SDS-PAGE indicated that only monomeric species were obtained for the C49A/C73A (lanes 1 and 2) and for the C65A/C86A (lanes 3 and 4) BPTI mutants i.e. no mixed disulfides were formed. Panel C) Analysis of the C49A/C73A mutants using immobilized metal affinity chromatography and subsequent analysis by reverse phase HPLC analysis. The elution positions of the native species containing 3 disulfides, the fully reduced species and the intermediate species containing only 1 or 2 disulfides are indicated. These results indicate that only a single two disulfide species is obtained.

Sequence listing

The sequence listing comprises the sequences:

- 20 SEQ ID NO:1 Endoplasmic oxidoreductin-1 from *Saccharomyces cerevisiae* (Swiss-Prot ID Q03103)
- SEQ ID NO:2 Human ERO1-like protein alpha (Swiss-Prot ID Q96HE7)
- SEQ ID NO:3 Human ERO1-like protein beta (Swiss-Prot ID Q86YB8)
- SEQ ID NO:4 Human ALR (Swiss-Prot ID P55789)
- 25 SEQ ID NO:5 *S. cerevisiae* mitochondrial FAD-linked sulfhydryl oxidase ERV1 (Swiss-Prot ID P27882)
- SEQ ID NO:6 *S. cerevisiae* FAD-linked sulfhydryl oxidase ERV2 (Swiss-Prot ID Q12284)
- SEQ ID NO:7 Human sulfhydryl oxidase 1 (QSOX; Swiss-Prot ID O00391)
- 30 SEQ ID NO:8 Human sulfhydryl oxidase 2 (Swiss-Prot ID Q6ZRP7)

SEQ ID NO:9 Vaccinia virus FAD-linked sulfhydryl oxidase E10 (Swiss-Prot ID P21050)

SEQ ID NO:10 *E.coli* (K12 strain) DsbC

SEQ ID NO:11 Human PDI

5 Detailed description of the invention

Definitions

“GSH” means reduced glutathione, a tripeptide called also γ -glutamylcysteinylglycine.

“GSSG” means oxidized glutathione. In GSSG two GSH's are linked by a disulfide bond.

- 10 “Oxidases” mean here enzymes that use molecular oxygen to oxidize a (bio)chemical compound.

“Thioredoxins” mean enzymes that catalyse the reduction of disulfide bonds, which are often found in protein substrates. Thioredoxin is reduced in turn by thioredoxin reductase.

- 15 “Glutaredoxins” mean enzymes that catalyse the removal of glutathione from glutathione-protein mixed disulfides using reduced glutathione and generating oxidized glutathione, which in turn is reduced by glutathione reductase.

- 20 “PDI” mean protein disulfide isomerase. PDI's are enzymes which catalyse thiol-disulfide exchange reactions in the endoplasmic reticulum (ER) of eukaryotes. They belong to the thioredoxin superfamily.

“Dsb” mean disulfide bond forming enzymes. Dsb's are enzymes which catalyse thiol-disulfide exchange reactions in the periplasm of prokaryotes. DsbA, DsbC and DsbG belong to the thioredoxin superfamily. DsbB and DsbD are not thiol-disulfide isomerases.

- 25 Thioredoxin superfamily members are for example thioredoxins, protein disulfide isomerases (PDI's) and disulfide bond forming enzymes (Dsb's).

The present invention relates to a method for producing a natively folded disulfide bond containing protein in a prokaryotic host. The method comprises that a prokaryo-

tic host cell is genetically engineered to express the protein of interest and a sulfhydryl oxidase in the cytoplasm of the host cell. The protein of interest is formed in a soluble form and contains disulfide bonds due to the presence of the sulfhydryl oxidase in the cytoplasm of said host cell. The present invention relates also to a prokaryotic host cell and a vector system for producing a protein of interest containing natively folded disulfide bonds.

A common prejudice in the field has been that the cytoplasm of most prokaryotes is reducing and therefore one cannot make disulfide bond containing proteins in the cytoplasm of a prokaryotic host unless the pathways for the reduction of disulfide bonds are disrupted or deleted, such as a $\Delta trxB / \Delta gor$ strain. No one has disclosed introducing an active catalyst of de novo disulfide bond formation into the cytoplasm of a prokaryotic host.

Pathways for disulfide bond formation in the endoplasmic reticulum, inter-membrane space of mitochondria and periplasm of prokaryotes are shown in **Figure 1**. Multiple pathways for dithiol oxidation to a disulfide in a folding protein exist. Direct oxidation by molecular oxygen, while widely used *in vitro*, is too slow to have physiological significance in native disulfide bond formation *in vivo*.

In the ER, PDI family members and GSSG can both introduce disulfide bonds into folding proteins, but both need to be reoxidised to complete the catalytic cycle. Oxidation of both PDI-family members and glutathione by molecular oxygen is too slow to have physiological significance in native disulfide bond formation *in vivo*. PDI is thought to be reoxidised by the sulfhydryl oxidase activity of Ero1-family members while the source of GSSG is under debate in the literature. The hydrogen peroxide made by Ero1 each catalytic cycle has the potential to oxidise dithiols in substrate proteins, to oxidise reduced glutathione to GSSG, to oxidise the active site of PDI family members and to form the regulatory disulfides in Ero1 and hence shut down peroxide production.

In the inter-membrane space of mitochondria of *Saccharomyces cerevisiae* Mia40 introduces disulfide bonds into folding proteins. To complete the catalytic cycle Mia40 is reoxidised by the sulfhydryl oxidase Erv1p. The direct formation of disulfide bonds in folding proteins or in Mia40 by molecular oxygen is too slow to have physiological significance in native disulfide bond formation *in vivo*. The potential

parallel role of glutathione in disulfide bond formation in the inter-membrane space of mitochondria is unknown.

In the periplasm DsbA introduces disulfide bonds into folding proteins. To complete the catalytic cycle DsbA is reoxidised by the transmembrane protein DsbB. Unlike
5 Ero1-family members or Erv1p, DsbB does not have sulfhydryl oxidase activity. The direct formation of disulfide bonds in folding proteins or in DsbB by molecular oxygen is too slow to have physiological significance in native disulfide bond formation *in vivo*. The potential parallel role of glutathione in disulfide bond formation in the periplasm is under debate in the literature.

10 Herein, by "protein of interest" are meant in particular proteins produced recombinantly in a foreign host. The proteins contain one or more disulfide bonds in their native state which are required to attain their native conformation. Many such proteins when expressed in a system in which disulfide bond formation is limited form insoluble inclusion bodies within the host. A protein of interest is here typically a
15 eukaryotic protein, usually a mammalian protein, in particular a human protein.

The protein may have in its native state multiple disulfide bonds. The method of the present invention is particularly suitable for producing a protein having in its native state two or more disulfide bonds.

By "a host" or "a foreign host" is meant here in particular a prokaryotic host. More
20 specifically the host can be a bacterial host, in particular a gram negative host, such as *Escherichia coli* (*E. coli*). In one preferred embodiment the host is *E. coli*.

According to this disclosure it is possible to produce a natively folded disulfide bond containing protein in a prokaryotic host. According to a preferred embodiment of the invention the protein may be recovered and optionally purified from the cultured host
25 cells. The protein may be lyophilized or formulated with a carrier or diluents, if needed.

The advantage of the present invention is that the protein of interest is formed in soluble form. Typically no denaturation and renaturation steps of the protein are needed. Furthermore, the protein is produced directly in biologically active form.

Commercially significant proteins which may be produced by using the present invention comprise for example insulin, blood coagulation factors, cytokines, chemokines, interferons, growth hormones and single chain antibodies.

5 In this disclosure as examples of such proteins are the luminal domain of human tissue factor, *E.coli* alkaline phosphatase and phytase, bovine pancreatic trypsin inhibitor (BPTI), human colony stimulating factor 3 (CSF3), bone morphogenic protein 4 (BMP4) and tissue plasminogen activator (t-PA). In the disclosure the ability to generate "soluble protein" or "insoluble inclusion bodies" is deduced from SDS-PAGE analysis of total and soluble fractions of a cell lysate. The generation of insoluble inclusion
10 bodies is a common occurrence when proteins that natively contain disulfide bonds are expressed in the cytoplasm of prokaryotic hosts. The formation of disulfide bonds allows protein folding to occur and hence allows the formation of soluble protein.

In the disclosure the "number of disulfide bonds" is deduced from the total number of cysteines in the protein and the number of cysteines free to react with iodoacetamide,
15 as determined from mass spectrometric analysis after treatment of the protein with iodoacetamide. The reaction of a thiol group with iodoacetamide adds 57 Da to the mass of the protein.

In the disclosure the "biological activity" of a protein is deduced by well known methods in the art appropriate for the individual proteins being assayed. The biological
20 activity or function of a protein reflects characteristics of the protein that result from the structure and conformational flexibility of the protein. These in turn are often dependent on the formation of native disulfide bonds. Hence biological activity, for example the ability of an enzyme to catalyze a specific enzymatic activity, is a measure of the attainment of the formation of native disulfide bonds within a protein.

25 The present invention provides a method for the production of proteins that require disulfide bond formation to reach their native biologically active conformation. The methods presented are particularly suitable for the expression of biologically active proteins that require the formation of multiple disulfide bonds. By multiple disulfide bonds is here meant two or more than two, typically more than three disulfide bonds.

30 By "a sequential disulfide bond" is meant a covalent linkage between the sulfur atoms of two cysteine residues in a protein which do not have an intervening cysteine residue in the primary sequence of the protein which is involved in another disulfide bond.

By “a non-sequential disulfide bond” is meant a covalent linkage between the sulfur atoms of two cysteine residues in a protein which have an intervening cysteine residue in the primary sequence of the protein which is involved in another disulfide bond.

The present invention is based on the use of pre- or co-expression of a sulfhydryl oxidase in the cytoplasm of prokaryotic bacteria, preferably of gram-negative bacteria such as *E.coli*, to generate disulfide bonds in folding proteins. As used herein, “sulfhydryl oxidases” refer to enzymes which use molecular oxygen to catalyze the reaction:



Since hydrogen peroxide (H_2O_2) is an oxidant that can be used to make disulfide bonds the overall reaction can also be written:



In addition, some sulfhydryl oxidases, such as Erv1, are reported also to be able to transfer electrons to the cytochrome respiratory chain in mitochondria instead of using molecular oxygen as the acceptor.

20 Sulfhydryl oxidases comprise proteins such as Ero1, Erv1, Erv2, QSox, ALR etc. These enzymes use FAD as a cofactor and contain one, or more, redox active disulfide bonds. They fall into two families:

i) the Ero family (Enzyme classification: EC 1.8.4.-) which are endoplasmic reticulum located sulfhydryl oxidases and which comprises, but is not limited to, the following exemplary members (see **Figure 3** for sequences): Endoplasmic oxidoreductin-1 from *Saccharomyces cerevisiae* (Swiss-Prot ID Q03103; for example as Genbank accession number CAA90553) (SEQ ID NO:1), human ERO1-like protein alpha (Swiss-Prot ID Q96HE7; for example as Genbank accession numbers AAF35260 or AAQ88828 or BAF85528 or AAH08674 or AAH12941) (SEQ ID NO:2), and human

ERO1-like protein beta (Swiss-Prot ID Q86YB8; for example as Genbank accession numbers AAF97547 or CAI23525 or CAI14420 or AAH32823 or AAH44573) (SEQ ID NO:3);

- ii) proteins which contain an ERV/ALR sulfhydryl oxidase domain (Enzyme classification: EC 1.8.3.2). These proteins are found in a variety of cellular compartments including mitochondria and the secretory pathway. They include, but are not limited to, the following exemplary members (see **Figure 4** for sequences): Human ALR (Swiss-Prot ID P55789; for example as Genbank accession numbers AAG38105) (SEQ ID NO:4), *Saccharomyces cerevisiae* mitochondrial FAD-linked sulfhydryl oxidase ERV1 (Swiss-Prot ID P27882; for example as Genbank accession numbers CAA97017) (SEQ ID NO:5), *Saccharomyces cerevisiae* FAD-linked sulfhydryl oxidase ERV2 (Swiss-Prot ID Q12284; for example as Genbank accession numbers CAA94987 or CAA92143) (SEQ ID NO:6), Human sulfhydryl oxidase 1 (QSOX; Swiss-Prot ID O00391; for example as Genbank accession numbers AAC09010 or AAQ89300 or CAI14838 or AAI00024) (SEQ ID NO:7), Human sulfhydryl oxidase 2 (Swiss-Prot ID Q6ZRP7; for example as Genbank accession numbers CAI16881 or CAM28352) (SEQ ID NO:8) and Vaccinia virus FAD-linked sulfhydryl oxidase E10 (Swiss-Prot ID P21050; for example as Genbank accession number AAA48051) (SEQ ID NO:9)

Within the scope of the present invention are proteins comprising any of the amino acid sequences of the above listed Genbank accession numbers or any of the amino acid sequences SEQ ID NO: 1 to 9, or a fragment of said sequences or a modified version of said sequences, which sequences still have sulfhydryl oxidase activity.

"A modified protein" refers to a protein wherein at one or more positions there have been amino acid insertions, deletions, or substitutions, either conservative or non-conservative, provided that such changes result in a protein still having sulfhydryl oxidase activity.

By "conservative substitutions" are meant here combinations, such as Val, Ile, Leu, Ala, Met; Asp, Glu; Asn, Gln; Ser, Thr, Gly, Ala; Lys, Arg, His; and Phe, Tyr, Trp. Preferred conservative substitutions comprise Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr, Lys, Arg; and Phe, Tyr.

A modified version (or variant) of the sulfhydryl oxidase comprises typically an amino acid sequence having at least 25 %, preferably at least 40 %, more preferably at least

50 %, still more preferably at least 60 %, still and still more preferably at least 70 %, more and more preferably at least 90 % identity, most preferably at least 95 %, or at least 98 % identity to any of the amino acid sequences of the above listed Genbank accession numbers or to any of the amino acid sequences SEQ ID NO: 1 to 9.

- 5 In addition to full length sulfhydryl oxidases, shorter fragments may be used as long as they retain sulfhydryl oxidase activity. A fragment refers to a protein having at one or more positions deletions. The fragment may comprise at least 30%, at least 40%, at least 50%, preferably at least 60 %, more preferably at least 70 %, still more preferably at least 80 %, more and more preferably at least 90 %, or at least 95 % of the
10 complete amino acid sequence of the proteins.

- By the term "identity" is here meant the identity between two amino acid sequences compared to each other from the first amino acid encoded by the corresponding gene to the last amino acid. Preferably the identity is measured by comparing the amino acid sequences without the sequences of any signal peptide the protein may have.
15 The identity of the full-length sequences may be measured for example by using sequence analysis software, for example BLAST software available from the National Library of Medicine.

- According to other preferred embodiments of the invention, the sulfhydryl oxidase of the invention is encoded by a nucleic acid sequence hybridizing under low or under
20 high stringent conditions to nucleic acid sequences encoding amino acid sequences of the above listed Genbank accession numbers or amino acid sequences SEQ ID NO: 1 to 9. By high stringency conditions are meant conditions as disclosed for example in Ausubel *et al.* Current Protocols in Molecular Biology, 1996, Wiley Sons, New York, N.Y.. High stringency hybridization conditions may comprise hybridization
25 at about 42 °C and about 50 % formamide, a first wash at about 65 °C, about 2 x SSC, and 1% SDS, followed by second wash at about 65 °C, and about 0.1 % SDS, 1xSSC. Lower stringency hybridization conditions may comprise hybridization at about 42 °C in the absence of formamide, a first wash at about 42 °C, about 6 x SSC, and 1 % SDS, followed by second wash at about 50 °C, about 6 x SSC and about 1
30 % SDS. Within the scope of the present invention are also nucleic acid sequences encoding the proteins comprising any of the amino acid sequences of the above listed Genbank accession numbers or any of the amino acid sequences SEQ ID NO: 1 to 9, or nucleic acid sequences encoding a fragment of said sequences or a modified ver-

sion of said sequences, which sequences still have sulfhydryl oxidase activity. Suitable nucleic acid sequences encoding proteins having sulfhydryl oxidase activity are publicly available and can be found in gene banks.

5 Within the scope of the present invention are any enzymes with sulfhydryl oxidase activity. In particular, within the scope of the present invention are sulfhydryl oxidases comprising ERV/ALR sulfhydryl oxidase domain, in particular Erv1.

10 Within the scope of the invention are sulfhydryl oxidases belonging to Ero1 family proteins, in particular Ero 1, which is an ER-resident protein having an N-terminal signal sequence, the signal sequence should be removed before cytoplasmic expression is possible. Ero1 has multiple disulfide bonds which are required for it to function. It cannot fold to an active state in the cytoplasm of wild-type prokaryotes. Therefore, modifications, such as making the host cell deficient in thioredoxin reductase and/or glutathione reductase activity are needed to obtain functional protein, or preferentially the pre-expression of another sulfhydryl oxidase such as Erv1p will generate func-
15 tional Ero1 in the cytoplasm. Ero1 works in the endoplasmic reticulum of eukaryotes, a compartment which has a significantly more oxidizing redox state than the cytoplasm. This arises due to the presence of oxidized glutathione. In the cytoplasm any oxidized glutathione made (by any route) is reduced by glutathione reductase (the *gor* gene product).

20 The present invention has been exemplified here by using the *Saccharomyces cerevisiae* enzyme Erv1p. Erv1p is a 189 amino acid FAD-dependent sulfhydryl oxidase that catalyzes disulfide bond formation. It is required for the import and folding of small cysteine-containing proteins in the mitochondrial inter-membrane space and is thought to form a redox cycle with Mia40.

25 In addition to wild type sulfhydryl oxidases, mutants may be used as long as they retain sulfhydryl oxidase activity.

In addition to wild type sulfhydryl oxidases, chimeric fusion proteins may be used as long as they retain sulfhydryl oxidase activity.

30 Sulfhydryl oxidase according to the present disclosure is produced in soluble and biologically active form in the cytoplasm of the prokaryotic host.

Furthermore, sulfhydryl oxidase according to the present disclosure is typically capable of functioning without a partner protein e.g. Erv1p can function without Mia40.

The present invention provides also a vector system comprising

- a vector encoding a protein of interest, or having an insertion site for a nucleic acid sequence encoding a protein of interest, and a sulfhydryl oxidase,

or

- a first vector encoding a protein of interest, or having an insertion site for a nucleic acid sequence encoding a protein of interest, and
- a second vector encoding a sulfhydryl oxidase.

- 10 According to the present disclosure the vector or vectors are constructed to be capable of expressing the protein of interest and the sulfhydryl oxidase in the cytoplasm of the prokaryotic host cell.

The system may further comprise a vector encoding a thiol-disulfide isomerase in the cytoplasm of the prokaryotic host cell.

- 15 The vector encoding a thiol-disulfide isomerase may be the same vector encoding the sulfhydryl oxidase and/or the protein of interest.

The vector system can be introduced to a suitable prokaryotic host cell and the host can be cultured to produce a protein of interest comprising natively folded disulfide bonds.

- 20 The vector or vectors comprise a nucleic acid sequence (a gene) encoding a protein of interest and a sulfhydryl oxidase, and possible a thiol-disulfide isomerase. Two or all of said nucleic acid sequences may be in the same vector or all of them may be in different vectors. The nucleic acid sequence may be linked to a nucleic acid sequence encoding a suitable part of a fusion protein, said nucleic acid construction encoding a fusion protein. The vector system furthermore comprises regulatory elements for multiplying and expressing the nucleic acid sequences in a prokaryotic host. Each mRNA produced from the vector(s) may be separately inducible. The vector system may comprise also selection markers.
- 25

Host cells comprising the expression vector encoding the sulfhydryl oxidase and a protein of interest are cultured to produce the protein of interest in a biologically active form. Any suitable expression vector may be used. It is preferable that the vector contains an appropriate selection marker. It is also preferable that the vector contains a system for inducing expression.

Methods for cloning the genes of interest into appropriate vectors and culturing prokaryotic organisms are well known in the art.

The construction of suitable vectors has been exemplified here in the examples which report the use of pET23 and pLysS plasmid derivatives. These have ampicillin and chloramphenicol selection markers, respectively. T7 and arabinose inducible expression systems were used.

Any suitable culture media may be used for the cultivation of the prokaryotic organisms. In the examples reported here Luria-Bertani Media (LB media) was used. M9 minimal media has also successfully been used.

The protein may be obtained from the cultured cells in a soluble form by routine cell lysis methods.

Cell lysis has been exemplified here in the examples by performing the addition of 0.1 mg/ml lysozyme to the resuspended cell pellet followed by freeze-thawing.

The protein of interest can be isolated from the cell lysate in substantially pure form by methods well known in the art and that are appropriate for the individual proteins and final application, for example column chromatography, polyacrylamide electrophoresis, or HPLC analysis. This can include the addition of a fusion tag to the protein of interest to aid purification.

Useful purification methods have been exemplified here in the examples where N-terminal hexa-histidine or N-terminal maltose binding protein (MBP) tags were used to facilitate purification using immobilized metal affinity chromatography or amylose resin, respectively.

"A substantially pure protein" means a preparation which is at least 60% by weight (dry weight) the protein of interest. Preferably the preparation is at least 75 %, more

preferably at least 80 %, still more preferably at least 90 %, still more preferably at least 95 %, most preferably at least 99 % by weight of the protein of interest.

In some applications the protein product comprising the protein of interest may be used with the cell culture without recovery, isolation and/or without purification. In other applications the protein product comprising the protein of interest may be recovered from the cell culture or cell medium or from host cells with or without purification. Furthermore, in some applications the protein product or purified protein may be diluted or concentrated, or lyophilized.

This invention has multiple possible alternative solutions. In one embodiment polycistronic vectors may be used i.e. vectors in which a single mRNA encodes the protein of interest plus the sulfhydryl oxidase and, where necessary, the thiol-disulfide isomerase (see **figure 5**). The order of the genes on polycistronic vectors does not affect the ability to co-expression of the proteins.

In another embodiment, the sulfhydryl oxidase, and where necessary the thiol-disulfide isomerase, are on separate vector from the protein of interest (see **figure 5**). The use of a two vector system allows for separate induction.

According to a preferred embodiment of the invention the vector or vectors is a plasmid or plasmids.

In one embodiment of the invention co-expression of a cytoplasmically expressed sulfhydryl oxidase with the protein of interest in wild type prokaryotic cells is undertaken. Disulfide bonds in the periplasm and in the ER are thought to form sequentially i.e. cysteine 1 is linked to cysteine 2, cysteine 3 to cysteine 4 etc. Hence proteins with a small number of sequential disulfide bonds should fold efficiently to their native state in the cytoplasm upon pre- or co-expression of a cytoplasmically expressed sulfhydryl oxidase. This is exemplified here with the efficient production of folded luminal domain of human tissue factor and *E.coli* alkaline phosphatase upon co-expression of the sulfhydryl oxidase (see examples 2 and 3). The sulfhydryl oxidase may be targeted to the cytoplasm by the removal of the N-terminal signal sequence where necessary.

In another embodiment of the invention, co-expression of a cytoplasmically expressed thiol-disulfide isomerase along with the sulfhydryl oxidase and the protein of interest is undertaken. For proteins with an increased number of sequential disulfide bonds

there is a greater probability that incorrect disulfide bonds will form. The formation of incorrect disulfide bonds or the formation of non-sequential native disulfide bonds in the ER and periplasm is known to require the subsequent action of a thiol-disulfide isomerase to attain the native disulfide state of the protein of interest. Co-expression of such an isomerase, for example DsbC or PDI targeted to the cytoplasm, and combined with the expression of the sulfhydryl oxidase and the protein of interest increases the yield of natively folded protein. This has been exemplified with *E.coli* phytase (see example 4) and kringle t-PA (see example 6) and other proteins. The isomerases may be targeted to the cytoplasm by the removal of their N-terminal signal sequence.

Within the scope of the present invention are any enzymes with thiol-disulfide isomerase activity.

Preferred PDI sequences (Enzyme classification EC 5.3.4.1) comprise human PDI or yeast Pdi1p sequences. Yeast Pdi1p sequences can be found for example as Genbank accession numbers CAA42373 or BAA00723 or CAA38402.

Human PDI family members are for example PDI (for example as Genbank accession numbers CAA28775 or AAC13652 (SEQ ID NO:11)), ERp57 (for example as Genbank accession numbers BAA03759 or AAC51518), PDlp (for example as Genbank accession numbers BAE38734 or AAK61223), ERp72 (for example as Genbank accession numbers AAA58460 or AAH00425), PDILT (for example as Genbank accession numbers BAC05068 or AAH42607), ERp27 (for example as Genbank accession numbers AAQ88900 or AAH30218), PDlr (for example as Genbank accession number BAA08451), ERp28 (for example as Genbank accession numbers CAA64397 or CAG46468), Erdj5 (for example as Genbank accession numbers AF038503 or AK027696), P5 (for example as Genbank accession numbers BAA08450 or AAH01312), ERp18 (for example as Genbank accession numbers AF543416 or CAI17031), ERp44 (for example as Genbank accession numbers CAC87611 or AAQ89407), ERp46 (for example as Genbank accession numbers AAQ89009 or BAC11526), TMX (for example as Genbank accession numbers AAQ89003 or AK075395), TMX2 (for example as Genbank accession numbers AAD27740 or AAH00666), TMX3 (for example as Genbank accession numbers BAG53687 or AAH93792), TMX4 (for example as Genbank accession numbers AAQ89363 or BAC11599), hAG-2 (for example as Genbank accession numbers AAC77358 or

AAQ89368) and hAG-3 (for example as Genbank accession numbers AAL55402 or AAH58284).

Sequences for DsbC (EC 5.3.4.1) can be found for example as Genbank accession numbers AAA83074 or AAC75931 (SEQ ID NO:10).

- 5 The thiol-disulfide isomerases having N-terminal signal sequence can be expressed in the cytoplasm by the removal of the N-terminal signal sequence.

In addition to mature thiol-disulfide isomerases, shorter fragments or modified forms or chimeric proteins may be used as long as they retain thiol-disulfide isomerase activity.

- 10 Enzymes having thiol-disulfide isomerase activity may comprise fragments or modifications of the above amino acid sequences.

"A modified protein" refers to a protein wherein at one or more positions there have been amino acid insertions, deletions, or substitutions, either conservative or non-conservative, provided that such changes result in a protein still having thiol-disulfide

- 15 isomerase activity.

A modified version (or variant) of the enzyme having thiol-disulfide isomerase activity comprises typically an amino acid sequence having at least 25 %, preferably at least 40 %, more preferably at least 50 %, still more preferably at least 60 %, still and still more preferably at least 70 %, more and more preferably at least 90 % identity, most

- 20 preferably at least 95 % or at least 98 % identity to any of the above mentioned amino acid sequences.

In addition to full length enzyme having thiol-disulfide isomerase activity, shorter fragments may be used as long as they retain thiol-disulfide isomerase activity. A fragment refers to a protein having at one or more positions deletions. The fragment

- 25 may comprise at least 30%, at least 40%, at least 50%, preferably at least 60 %, more preferably at least 70 %, still more preferably at least 80 %, more and more preferably at least 90 %, or at least 95 % of the complete amino acid sequence of the proteins.

According to other preferred embodiments of the invention, the enzyme having thiol-disulfide isomerase activity of the invention is encoded by a nucleic acid sequence

- 30

hybridizing under low or under high stringent conditions to nucleic acid sequences encoding the above mentioned amino acid sequences. By high stringency conditions are meant conditions as disclosed for example in Ausubel *et al.* Current Protocols in Molecular Biology, 1996, Wiley Sons, New York, N.Y. High stringency hybridization conditions may comprise hybridization at about 42 °C and about 50 % formamide, a first wash at about 65 °C, about 2 x SSC, and 1% SDS, followed by second wash at about 65 °C, and about 0.1 % SDS, 1xSSC. Lower stringency hybridization conditions may comprise hybridization at about 42 °C in the absence of formamide, a first wash at about 42 °C, about 6 x SSC, and 1 % SDS, followed by second wash at about 50 °C, about 6 x SSC and about 1 % SDS.

Within the scope of the present invention are also nucleic acid sequences encoding the proteins comprising any of the amino acid sequences of the above gene bank accession numbers or the sequences SEQ ID NO:10 or SEQ ID NO:11 having thiol-disulfide isomerase activity, or nucleic acid sequences encoding a fragment of said sequences or a modified version of said sequences, which sequences still have thiol-disulfide isomerase activity. Suitable nucleic acid sequences encoding proteins having thiol-disulfide isomerase activity are publicly available and can be found in gene banks.

In another embodiment of the invention, the protein of interest is expressed as a fusion protein. Disulfide bond formation is the rate limiting step in protein folding *in vitro* and *in vivo*. If the folding protein misfolds and aggregates (forms inclusion bodies) before the sulfhydryl oxidase and/or thiol-disulfide isomerase can act on it then low yields of correctly folded protein will be obtained. The generation of a fusion protein between a well expressed and quickly folding protein and the protein of interest is known to often inhibit inclusion body formation. Well known in the art examples include maltose binding protein (MBP), glutathione S-transferase (GST) and thioredoxin (Trx). In the examples included here we used MBP. The use of this fusion protein technology combined with co-expression of the sulfhydryl oxidase increased the yield of soluble protein for human CSF3 and BMP4 (see example 5). When this technology was combined with co-expression of a thiol-disulfide isomerase as well it increased the soluble yields of kringle t-PA (see example 6) and other proteins.

In another embodiment of the invention, mutant prokaryotic cells are used. Usually the cytoplasm of *E.coli* is reducing and any disulfides that are formed can be reduced

by thioredoxin / thioredoxin reductase and/or the glutathione / glutaredoxin / glutathione reductase systems while the protein is folding. To inhibit this, the expression of the sulfhydryl oxidase can be combined with knock-outs in these two pathways, for example the commercial rosetta-gami strain. When expression of the sulfhydryl oxidase was combined with the rosetta-gami strain background it increased the yields of active kringle tPA (see example 6) but decreased the yields of active alkaline phosphatase (see example 3) and phytase (see example 4).

In another embodiment of the system the sulfhydryl oxidase, and where appropriate a thiol-disulfide isomerase, are expressed prior to expression of the protein of interest. The kinetic partitioning between productive folding and the formation of inclusion bodies is known to depend on the degree to which protein aggregates are already present in the system. Upon co-expression of a sulfhydryl oxidase with the protein of interest, the initial production of the protein of interest is effective in the absence of sulfhydryl oxidase, since both are being co-expressed. This increases the chances of initial aggregate formation which in turn becomes an auto-catalytic process. The pre-synthesis of the sulfhydryl oxidase, and thiol-disulfide isomerase if required, prior to the induction of synthesis of the protein of interest circumvents this problem and would be expected to increase the soluble biologically-active yields of the protein of interest. Suitable vectors based on pLysS with an arabinose induction system to enable this were constructed (see **Figure 5**). This system is able to fold a range of proteins and increases the yield of active protein for example see kringle t-PA (example 6).

According to the present disclosure

- The protein production system according to this disclosure increases production of active protein in prokaryotic hosts compared to wild type strains. The increase is over 100-fold in a wild-type *E.coli* background. The present system also gives circa 3 fold higher activity of protein of interest than the prior system (rosetta-gami strain). This has been here exemplified by expressing alkaline phosphatase, tPA and phytase in this system and in the commercial rosetta-gami strain.
- The rosetta-gami strain is genetically unstable and grows much slower than the comparable wild-type strains. The use of wild type strains modified according to the present disclosure is more advantageous. Although, for some proteins the use of a

strain being deficient in thioredoxin reductase or glutathione reductase activity and being modified according to the present disclosure, may be of advantage.

- In addition to sulfhydryl oxidase the expression of disulfide isomerase may be of advantage, in particular if the protein of interest comprises non-sequential disulfide bonds or a large number of sequential disulfide bonds.

- The more proteins that are required to be co-expressed, the smaller the amount of cellular resources available to make the protein of interest i.e. the system must always be kept to the simplest possible.

- The more disulfide bonds the target protein has the higher the level of required expression of the sulfhydryl oxidase and the disulfide isomerase. However, too high a level of expression of either factor will result in a reduction in the levels of expression of the target protein (see above).

- Target proteins with non-optimal codon-usage for *E.coli* production may require the co-expression of rare tRNA's for optimal production e.g. the use of the pLysS RARE plasmid (found in rosetta-gami along with the *gor* and *trxB* knockouts).

- Target proteins that naturally have a pro-sequence generally may require the use of a fusion protein to keep them soluble during folding.

Various embodiments of the invention are described below with the aid of the following numbered clauses 1-19:

1. A method for producing a protein of interest containing one or more disulfide bonds in its native state, which comprises that a prokaryotic host cell is genetically engineered to express the protein of interest and a sulfhydryl oxidase in the cytoplasm of the host cell, said protein of interest being formed in a soluble form and containing disulfide bonds due to the presence of the sulfhydryl oxidase in the cytoplasm of said host cell.

2. The method according to clause 1, wherein the protein product comprising the protein of interest is recovered from the cell culture or from the host cells and optionally purified.

3. The method according to clause 1 or 2, wherein the protein of interest is co-expressed with the sulfhydryl oxidase.

4. The method according to any one of clauses 1 to 3, wherein the sulfhydryl oxidase is expressed prior to the protein of interest.
5. The method according to any one of the preceding clauses, wherein the host cells are engineered to further express a thiol-disulfide isomerase in the cytoplasm.
- 5 6. A prokaryotic host cell for producing a protein of interest containing natively folded disulfide bonds, which comprises that the host cell is genetically engineered to express a sulfhydryl oxidase and a protein of interest in the cytoplasm of the host cell.
7. The method or the host cell according to any one of the preceding clauses, wherein the sulfhydryl oxidase uses FAD as a cofactor and contains one, or more, redox active disulfide bonds.
- 10 8. The method or the host cell according to any one of the preceding clauses, wherein the sulfhydryl oxidase belongs to sulfhydryl oxidases comprising ERV/ALR sulfhydryl oxidase domain.
9. The method or the host cell according to any one of the preceding clauses, wherein the sulfhydryl oxidase belongs to ERO family of sulfhydryl oxidases.
- 15 10. The method or the host cell according to any one of the preceding clauses, wherein the host cell further expresses a thiol-disulfide isomerase in the cytoplasm.
11. The method or the host cell according to clause 10, wherein the disulfide isomerase is PDI, or DsbC.
- 20 12. The method or the host cell according to any one of the preceding clauses, wherein the prokaryotic host is a gram negative bacterium, such as *E. coli*.
13. The method or the host cell according to any one of the preceding clauses, wherein the protein of interest is produced as a fusion protein.
14. The method or the host cell according to any one of the preceding clauses, wherein the sulfhydryl oxidase and the disulfide isomerase are induced either separately or together with the protein of interest.
- 25 15. The method or the host cell according to any one of the preceding clauses, wherein the host cell is deficient in thioredoxin reductase and/or glutathione reductase activity.

16. A vector system for a prokaryotic host cell for producing a protein of interest containing natively folded disulfide bonds, which comprises

- a vector encoding a protein of interest, or having a site for a nucleic acid sequence encoding a protein of interest, and a sulfhydryl oxidase,

5 or

- a first vector encoding a protein of interest, or having a site for a nucleic acid sequence encoding a protein of interest, and

- a second vector encoding a sulfhydryl oxidase,

10 wherein said vector or vectors are constructed to be capable of expressing the protein of interest and the sulfhydryl oxidase in the cytoplasm of a prokaryotic host cell.

17. The vector system according to clause 16 for producing a protein of interest, wherein the system further comprises a vector encoding a thiol-disulfide isomerase, said vector being the same or different vector as the vectors in the vector system, and said vector being constructed to be capable of expressing a thiol-disulfide isomerase
15 in the cytoplasm of the prokaryotic host.

18. The vector system according to clause 16 or 17, wherein the sulfhydryl oxidase and thiol-disulfide isomerase are induced either separately or together with the protein of interest.

19. A prokaryotic host cell system, which comprises the vector system according to
20 any one of clauses 16 to 18.

The following non-limiting examples illustrate the invention.

Examples

Example 1

25 Plasmid constructs used in the protein expressions were generated as part of pre-studies (see Figure 5). The original constructs were polycistronic with multiple genes encoded by a single mRNA driven by a T7 promoter system (IPTG inducible) in a modified version of pET23. These included the protein of interest alone

(A), this plus the sulfhydryl oxidase (B) or plus the sulfhydryl oxidase and either PDI or DsbC as a thiol-disulfide isomerase (C).

Subsequent constructs had the sulfhydryl oxidase (D) or the sulfhydryl oxidase plus the thiol-disulfide isomerase (E) on an arabinose promoter as part of a modified version of pLysS. Constructs D and E are fully compatible with construct A and allow easy inter-conversion of the protein of interest in the system. In addition, constructs D and E allow pre-expression or co-expression of the sulfhydryl oxidase, or the sulfhydryl oxidase and thiol-disulfide isomerase, and the protein of interest.

10 Example 2

Efficient production of the luminal domain of human tissue factor

Tissue factor (TF), also known as thromboplastin factor III, is a protein involved in the coagulation of blood. It is a transmembrane protein whose luminal domain contains two sequential disulfide bonds.

15 pVD81, is a derivative of pET23a which has cloned into the multi-cloning site a gene encoding for an N-terminal hexa-histidine tag (MHHHHHHM) followed in frame with the luminal domain of human tissue factor (sTF) as represented by the fragment Ser 33 -Glu 251 of the full length protein. This vector expresses sTF upon induction with IPTG.

20 pVD77 is a derivative of pVD81, in which the gene for the sulfhydryl oxidase Erv1p from *S.cerevisiae* has been cloned (Met 1 – Glu 189) (SEQ ID NO: 5) after the gene for sTF (with suitable ribosome binding sites to initiate translation of both; see **Figure 5**). This polycistronic vector co-expresses sTF and Erv1p upon induction with IPTG.

25 *E.coli* strains transformed with these expression vectors were streaked out from glycerol stocks stored at -70°C onto agar plates containing suitable antibiotics to allow for selection. The next day one colony from these plates were used to inoculate 5 mls of LB media, containing suitable antibiotics, and grown overnight at 30°C, 200 rpm. This overnight culture was used to seed a 50 mls culture of LB containing suitable antibiotics in a 250 ml flask to an optical density of 0.05 at 600 nm
30 (OD₆₀₀). This culture was grown at 25°C, 200 rpm until the OD₆₀₀ reached 0.4 at

which point protein production was induced by the addition of 1 mM IPTG. The cells were then grown for 4 hours at 25°C, 200 rpm and the final OD₆₀₀ measured. The cells were collected by centrifugation and resuspended to an OD₆₀₀ equivalent of 10 in 20 mM sodium phosphate pH 7.4, 20 µg/ml DNase, 0.1 mg/ml egg white
5 lysozyme, and frozen. Cells were lysed by freeze-thawing. Where appropriate, the insoluble fraction was removed by centrifugation and the soluble fraction removed quickly to a new container.

When sTF is expressed alone in the cytoplasm of the *E.coli* strain BL21 (DE3) pLysS the protein is unable to fold correctly and forms insoluble inclusion bodies.
10 When a sulfhydryl oxidase is co-expressed at 25°C with sTF in the cytoplasm of the same *E.coli* strain, sTF is made solubly in high yields. **Figure 6** shows representative SDS-PAGE from expression of sTF with co-expression with a sulfhydryl oxidase in LB media at 25°C. Lane 2 shows the total *E.coli* lysate, lane 1 shows the soluble fraction. The position of the luminal domain of tissue factor is marked
15 with an arrow.

Purification of this protein via an N-terminal hexa-histidine tag followed by treatment with iodoacetamide and mass spectrometry analysis of the purified protein revealed the existence of no free thiol groups implying that two disulfide bonds had been formed in this system.

20 **Example 3**

Efficient production of *E.coli* alkaline phosphatase.

Alkaline phosphatase is a hydrolase which removes phosphate groups from many types of molecules. The bacterial enzyme folds in the periplasm. It has two sequential disulfide bonds whose formation is essential for activity. Since it is easily assayed
25 bacterial alkaline phosphatase is widely used as a model protein for disulfide bond formation *in vivo*.

pVD80, is a derivative of pET23a which has cloned into the multi-cloning site a gene encoding for an N-terminal hexa-histidine tag (MHHHHHHM) followed in frame with the mature form of *E.coli* alkaline phosphatase (PhoA) as represented by the frag-
30 ment Arg 22 -.Lys 471 of the full length protein. This vector expresses alkaline phosphatase upon induction with IPTG.

pVD82 is a derivative of pVD80, in which the gene for the sulfhydryl oxidase Erv1p from *S.cerevisiae* has been cloned (Met 1 – Glu 189) (SEQ ID NO: 5) after the gene for *E.coli* alkaline phosphatase (with suitable ribosome binding sites to initiate translation of both; see **Figure 5**). This polycistronic vector co-expresses *E.coli* alkaline phosphatase and Erv1p upon induction with IPTG.

E.coli strains with these expression vectors were streaked out from glycerol stocks stored at -70°C onto agar plates containing suitable antibiotics to allow for selection. The next day one colony from these plates were used to inoculate 5 mls of LB media, containing suitable antibiotics, and grown overnight at 30°C, 200 rpm. This overnight culture was used to seed a 50 mls culture of LB containing suitable antibiotics in a 250 ml flask to an optical density of 0.05 at 600 nm (OD₆₀₀). This culture was grown at 30°C, 200 rpm until the OD₆₀₀ reached 0.4 at which point protein production was induced by the addition of 1 mM IPTG. The cells were then grown for 4 hours at 30°C, 200 rpm and the final OD₆₀₀ measured. The cells were collected by centrifugation and resuspended to an OD₆₀₀ equivalent of 10 in 20 mM sodium phosphate pH 7.4, 20 µg/ml DNase, 0.1 mg/ml egg white lysozyme and frozen. Cells were lysed by freeze-thawing. Where appropriate, the insoluble fraction was removed by centrifugation and the soluble fraction removed quickly to a new container.

To monitor the production of active alkaline phosphatase we used the standard 4-nitrophenylphosphate assay at pH 8.0. This assay revealed that expression of alkaline phosphatase in the cytoplasm of *E.coli* resulted in the production of minimal active protein. In contrast co-expression of a sulfhydryl oxidase resulted in significant levels of active protein (see **Figure 7**).

Figure 7 shows measurements for the activity of *E.coli* alkaline phosphatase expressed in the cytoplasm of *E.coli* in LB media at 30°C. 4-nitrophenylphosphate used as the substrate, pH of the reaction was 8.0 and the values are cited as relative activity (%) compared with the most active system which is BL21 (DE3) pLysSRARE + expression of the sulfhydryl oxidase. SO indicates co-expression of the sulfhydryl oxidase. Strains used were BL = BL21 (DE3) pLysRARE and RG = rosetta-gami.

Previously it had been shown that active alkaline phosphatase can be made in strains such as rosetta-gami in which the reducing pathways have been disrupted. Production of alkaline phosphatase in rosetta-gami resulted in the production of active pro-

tein, but circa three-fold less activity was seen than for the co-expression of the sulfhydryl oxidase in wild type *E.coli*.

Our system increases production of active protein over 100-fold in a wild-type *E.coli* background and gives circa 3 fold higher alkaline phosphatase activity than the commercial rosetta-gami strain.

Example 4

Efficient production of *E.coli* phytase

Phytase (AppA) is an *E.coli* protein with similar activity to alkaline phosphatase except that it has optimal activity under acidic pH values (hence one of its alternative names is acid phosphatase). The protein contains four disulfide bonds, one of which is non-sequential. As it contains a non-sequential disulfide bond it is used as a model protein for disulfide bond formation which requires isomerisation. It is also used as an important animal feed adduct.

pVD96, is a derivative of pET23a which has cloned into the multi-cloning site a gene encoding for an N-terminal hexa-histidine tag (MHHHHHMH) followed in frame with the mature form of *E.coli* Phytase (AppA) as represented by the fragment Gln 23 - Leu 432 of the full length protein. This vector expresses phytase upon induction with IPTG.

pFH231 is a derivative of pVD96, in which the gene for the sulfhydryl oxidase Erv1p from *S.cerevisiae* has been cloned (Met 1 – Glu 189) (SEQ ID NO: 5) after the gene for phytase (with suitable ribosome binding sites to initiate translation of both; see **Figure 5**). This polycistronic vector co-expresses phytase and Erv1p upon induction with IPTG.

pFH244 is a derivative of pVD96, in which the gene for the mature form of DsbC from *E.coli* has been cloned (Asp 21 – Lys 236) (SEQ ID NO: 10) after the gene for phytase (with suitable ribosome binding sites to initiate translation of both; see **Figure 5**). This polycistronic vector co-expresses both phytase and mature DsbC upon induction with IPTG.

pFH245 is a derivative of pVD96, in which the gene for the mature form of PDI from *H.sapiens* has been cloned (Asp 18 – Leu 508) (SEQ ID NO: 11) after the gene for

phytase (with suitable ribosome binding sites to initiate translation of both; see **Figure 5**). This polycistronic vector co-expresses phytase and mature PDI upon induction with IPTG.

5 pFH233 is a derivative of pVD96, in which the gene for the sulfhydryl oxidase Erv1p from *S.cerevisiae* (Met 1 – Glu 189) (SEQ ID NO:5) and the gene for the mature form of DsbC from *E.coli* (Asp 21 – Lys 236) (SEQ ID NO:10) have been cloned after the gene for phytase (with suitable ribosome binding sites to initiate translation of both; see **Figure 5**). This polycistronic vector co-expresses phytase, Erv1p and DsbC upon induction with IPTG.

10 pFH232 is a derivative of pVD96, in which the gene for the sulfhydryl oxidase Erv1p from *S.cerevisiae* (Met 1 – Glu 189) (SEQ ID NO:5) and the gene for the mature form of PDI from *H.sapiens* (Asp 18 – Leu 508) (SEQ ID NO: 11) have been cloned after the gene for phytase (with suitable ribosome binding sites to initiate translation of all; see **Figure 5**). This polycistronic vector co-expresses phytase, Erv1p and PDI upon
15 induction with IPTG.

E.coli strains with these expression vectors were streaked out from glycerol stocks stored at -70°C onto agar plates containing suitable antibiotics to allow for selection. The next day one colony from these plates were used to inoculate 5 mls of LB media, containing suitable antibiotics, and grown overnight at 30°C, 200 rpm. This overnight
20 culture was used to seed a 50 mls culture of LB containing suitable antibiotics in a 250 ml flask to an optical density of 0.05 at 600 nm (OD₆₀₀). This culture was grown at 30°C, 200 rpm until the OD₆₀₀ reached 0.4 at which point protein production was induced by the addition of 1 mM IPTG. The cells were then grown for 4 hours at 30°C, 200 rpm and the final OD₆₀₀ measured. The cells were collected by centrifugation and
25 resuspended to an OD₆₀₀ equivalent of 10 in 20 mM sodium phosphate pH 7.4, 20 µg/ml DNase, 0.1 mg/ml egg white lysozyme and frozen. Cells were lysed by freeze-thawing. Where appropriate, the insoluble fraction was removed by centrifugation and the soluble fraction removed quickly to a new container.

To monitor the activity of phytase an assay similar to that employed for alkaline phosphatase was used using 4-nitrophenylphosphate as the substrate (see example 3)
30 except that the pH of the reaction was held at pH 2.5, the optimal for phytase activity.

Figure 8 shows measurements for the activity of *E.coli* phytase (AppA) expressed in the cytoplasm of *E.coli* strains in LB media at 30°C. Absorbance measured after quenching the reaction with sodium hydroxide (NaOH) and the values are cited as relative activity (%) compared with the most active system which is BL21 (DE3) pLysSRARE + expression of the sulfhydryl oxidase and DsbC. The background activity is subtracted. Strains used were BL = BL21 (DE3) pLysSRARE and RG = rosetta-gami. SO indicates co-expression of the sulfhydryl oxidase, +D indicates co-expression of DsbC, +P indicates co-expression of PDI.

Over-expression of phytase in the cytoplasm of *E.coli* results in activity levels comparable with samples in which no phytase is expressed. In contrast co-expression of a sulfhydryl oxidase results in increased phytase activity (sample "BL+SO"). Since phytase contains a non-sequential disulfide bond co-expression of a thiol-disulfide isomerase is required for optimal folding. Note however that in the wild type background expression co-expression of only a thiol-disulfide isomerase does not significantly increase the activity of phytase (samples "BL + D" and "BL + P") they only aid so long as a sulfhydryl oxidase is present (samples "BL + SO + D" and "BL + SO + P"; see **Figure 8**).

The present invention increases production of active protein over 100-fold in a wild-type *E.coli* background and gives more activity than the current commercial rosetta-gami system.

Example 5

Efficient production of soluble human CSF3 and human BMP4

Colony stimulating factors (CSF) are hormones that cause cells to proliferate and differentiate. CSF3 stimulates the production of granulocytes and stem cells and is also known as granulocyte colony-stimulating factor. It is used in oncological and hematological applications and is made commercially under different names, including Neupogen and Religast. CSF3 contains two sequential disulfide bonds and one free thiol group near the N-terminal of the protein which probably results in non-native disulfide bond formation.

Bone morphogenic proteins (BMP) are a group of cytokines which induce the formation of bone and cartilage. BMP4 regulates the formation of teeth, limbs and bones

and plays a role in fracture repair. It contains three non-sequential intra-molecular disulfide bonds along with one inter-molecular disulfide bond.

pVD109 is a polycistronic derivative of pET23a which has a gene encoding for a cytoplasmically targeted MBP tag (as represented by the fragment Lys 27 – Thr 392 plus a linker encoding the amino acid sequence NSSSNNNNHM) followed in frame with the mature form of mature human CSF3 as represented by the fragment Ala 30 – Pro 207 of the full length protein. After this MBP-CSF3 fusion gene product the gene for the sulfhydryl oxidase Erv1p from *S.cerevisiae* has been cloned (Met 1 – Glu 189) (SEQ ID NO: 5) , with suitable ribosome binding sites to initiate translation of both; see **Figure 5**. This polycistronic vector co-expresses MBP-CSF3 and Erv1p upon induction with IPTG.

pVD110 is a derivative of pVD109, in which the gene for the mature form of DsbC from *E.coli* (Asp 21 – Lys 236) (SEQ ID NO: 10) has been cloned before the gene for MBP-CSF3 (with suitable ribosome binding sites to initiate translation of all; see **Figure 5**). This polycistronic vector co-expresses MBP-CSF3, Erv1p and DsbC upon induction with IPTG.

pVD111 is a derivative of pVD109, in which the gene for the mature form of PDI from *H.sapiens* (Asp 18 – Leu 508) (SEQ ID NO: 11) has been cloned before the gene for MBP-CSF3 (with suitable ribosome binding sites to initiate translation of all; see **Figure 5**). This polycistronic vector co-expresses MBP-CSF3, Erv1p and PDI upon induction with IPTG.

pVD112 is a polycistronic derivative of pET23a which has a gene encoding for a cytoplasmically targeted MBP tag (as represented by the fragment Lys 27 – Thr 392 plus a linker encoding the amino acid sequence NSSSNNNNHM) followed in frame with the mature form of mature human BMP4 as represented by the fragment Pro 294 – Arg 408 of the full length protein. After this MBP-CSF3 fusion gene product the gene for the sulfhydryl oxidase Erv1p from *S.cerevisiae* has been cloned (Met 1 – Glu 189), with suitable ribosome binding sites to initiate translation of both; see **Figure 5**. This polycistronic vector co-expresses MBP-BMP4 and Erv1p upon induction with IPTG.

pVD114 is a derivative of pVD112, in which the gene for the mature form of DsbC from *E.coli* (Asp 21 – Lys 236) (SEQ ID NO: 10) have been cloned before the gene

for MBP-BMP4 (with suitable ribosome binding sites to initiate translation of all; see **Figure 5**). This polycistronic vector co-expresses MBP-BMP4, Erv1p and DsbC upon induction with IPTG.

5 pVD113 is a derivative of pVD112, in which the gene for the mature form of PDI from *H.sapiens* (Asp 18 – Leu 508) (SEQ ID NO: 11) have been cloned before the gene for MBP-BMP4 (with suitable ribosome binding sites to initiate translation of all; see **Figure 5**). This polycistronic vector co-expresses MBP-BMP4, Erv1p and PDI upon induction with IPTG.

10 *E.coli* strains with these expression vectors were streaked out from glycerol stocks stored at -70°C onto agar plates containing suitable antibiotics to allow for selection. The next day one colony from these plates were used to inoculate 5 mls of LB media, containing suitable antibiotics, and grown overnight at 30°C, 200 rpm. This overnight culture was used to seed a 50 mls culture of LB containing suitable antibiotics in a 250 ml flask to an optical density of 0.05 at 600 nm (OD₆₀₀). This culture was grown at 15 30°C, 200 rpm until the OD₆₀₀ reached 0.4 at which point protein production was induced by the addition of 1 mM IPTG. The cells were then grown for 4 hours at 30°C, 200 rpm and the final OD₆₀₀ measured. The cells were collected by centrifugation and resuspended to an OD₆₀₀ equivalent of 10 in 20 mM sodium phosphate pH 7.4, 20 µg/ml DNase, 0.1 mg/ml egg white lysozyme and frozen. Cells were lysed by freeze- 20 thawing. Where appropriate, the insoluble fraction was removed by centrifugation and the soluble fraction removed quickly to a new container.

Figure 9 shows SDS-PAGE from expression of the human CSF3 (lanes 2-7) and BMP4 (lanes 8-13) as fusion proteins with MBP in the cytoplasm of *E.coli* BL21 (DE3) pLysS with co-expression with a sulfhydryl oxidase in LB media at 30°C and co- 25 expression of PDI (lanes 4,5,10,11) or DsbC (lanes 6,7,12,13). Even numbered lanes show the total *E.coli* lysates, odd numbered lanes show the soluble fractions (except lane 1 which shows molecular weight markers). The position of the MBP fusion proteins CSF3 and BMP4 are marked with arrows. In *E.coli* BL21 (DE3) pLysS with no co-expression of a sulfhydryl oxidase no soluble protein expression can be observed 30 for either protein but our system allows soluble expression of both.

The expression of CSF3 or BMP4 in the cytoplasm of *E.coli* BL21 (DE3) pLysS results in the formation of inclusion bodies. However, the expression of either protein in

fusion with MBP along with co-expression of a sulfhydryl oxidase and the isomerase PDI or DsbC results in production of soluble protein (see **Figure 9**).

Example 6

Efficient production of active kringle tPA

- 5 Tissue plasminogen activator (tPA) is a protease that converts plasminogen to plasmin, the major enzyme involved in the breakdown of blood clots. It is used medically to treat pulmonary embolism, myocardial infarction and stroke. It is a large protein that contains 16 non-sequential and one sequential disulfide bond and in addition has one free thiol group in the native structure. It is available under the names Activase and Retavase. Retavase is a fragment of tPA which only contains the kringle 2 and protease domains. As such it contains only 9 disulfide bonds.

pKEHS1156, is a derivative of pET23a which has cloned into the multi-cloning site a gene encoding for an N-terminal hexa-histidine tag (MHHHHHHM) followed in frame with the form of Human tPA which contains the kringle 2 and protease domains as represented by the fragment Gly 211 - Pro 562 of the full length protein. This vector expresses Kringle-tPA upon induction with IPTG.

pKEHS1165 is a derivative of pKEHS1156, in which the gene for the sulfhydryl oxidase Erv1p from *S.cerevisiae* has been cloned (Met 1 – Glu 189) after the gene for Kringle-tPA (with suitable ribosome binding sites to initiate translation of both; see **Figure 5**). This polycistronic vector co-expresses Kringle-tPA and Erv1p upon induction with IPTG.

pFH219 is a derivative of pKEHS1156, in which the gene for the sulfhydryl oxidase Erv1p from *S.cerevisiae* (Met 1 – Glu 189) (SEQ ID NO: 5) and the gene for the mature form of DsbC from *E.coli* (Asp 21 – Lys 236) (SEQ ID NO: 10) have been cloned after the gene for Kringle-tPA (with suitable ribosome binding sites to initiate translation of all; see **Figure 5**). This polycistronic vector co-expresses Kringle-tPA, Erv1p and DsbC upon induction with IPTG.

pVD103, is a derivative of pET23a which has cloned into the multi-cloning site a gene encoding for a cytoplasmically targeted N-terminal MBP tag (as represented by the fragment Lys 27 – Thr 392 plus a linker encoding the amino acid sequence NSSSNNNNHM) followed in frame with the form of Human tPA which contains the

kringle 2 and protease domains as represented by the fragment Gly 211 - Pro 562 of the full length protein. This vector expresses the MBP-Kringle-tPA fusion protein upon induction with IPTG.

5 pVD100 is a derivative of pVD103, in which the gene for the sulfhydryl oxidase Erv1p from *S.cerevisiae* has been cloned (Met 1 – Glu 189) after the gene for MBP-Kringle-tPA (with suitable ribosome binding sites to initiate translation of both; see **Figure 5**). This polycistronic vector co-expresses MBP-Kringle-tPA and Erv1p upon induction with IPTG.

10 pVD105 is a derivative of pVD103, in which the gene for the mature form of DsbC from *E.coli* has been cloned (Asp 21 – Lys 236) (SEQ ID NO: 10) before the gene for MBP-Kringle-tPA (with suitable ribosome binding sites to initiate translation of both; see **Figure 5**). This polycistronic vector co-expresses MBP-Kringle-tPA and mature DsbC upon induction with IPTG.

15 pVD104 is a derivative of pVD103, in which the gene for the mature form of PDI from *H.sapiens* has been cloned (Asp 18 – Leu 508) (SEQ ID NO: 11) before the gene for MBP-Kringle-tPA (with suitable ribosome binding sites to initiate translation of both; see **Figure 5**). This polycistronic vector co-expresses MBP-Kringle-tPA and mature PDI upon induction with IPTG.

20 pVD102 is a derivative of pVD100, in which the gene for the mature form of DsbC from *E.coli* (Asp 21 – Lys 236) (SEQ ID NO: 10) has been cloned before the genes for MBP-Kringle-tPA and Erv1p (with suitable ribosome binding sites to initiate translation of all; see **Figure 5**). This polycistronic vector co-expresses MBP-Kringle-tPA, Erv1p and DsbC upon induction with IPTG.

25 pVD101 is a derivative of pVD100, in which the gene for the mature form of PDI from *H.sapiens* (Asp 18 – Leu 508) (SEQ ID NO: 11) have been cloned before the genes for MBP-Kringle-tPA and Erv1p (with suitable ribosome binding sites to initiate translation of all; see **Figure 5**). This polycistronic vector co-expresses MBP-Kringle-tPA, Erv1p and PDI upon induction with IPTG.

30 pKEHS1208 is a derivative of pLysS in which the gene for sulfhydryl oxidase Erv1p from *S.cerevisiae* (Met 1 – Glu 189) (SEQ ID NO: 5) and the gene for the mature form of PDI from *H.sapiens* (Asp 18 – Leu 508) (SEQ ID NO: 11) have been cloned under

an arabinose inducible promoter/terminator system. This vector allows pre-expression of Erv1p and PDI upon induction with arabinose.

pKEHS1209 is a derivative of pLysS in which the gene for sulfhydryl oxidase Erv1p from *S.cerevisiae* (Met 1 – Glu 189) (SEQ ID NO: 5) and the gene for the mature form of DsbC from *E.coli* (Asp 21 – Lys 236) (SEQ ID NO: 10) have been cloned under an arabinose inducible promoter/terminator system. This vector allows pre-expression of Erv1p and DsbC upon induction with arabinose.

E.coli strains with these expression vectors (excluding strains which contain pKEHS1208 or pKEHS1209) were streaked out from glycerol stocks stored at -70°C onto agar plates containing suitable antibiotics to allow for selection. The next day one colony from these plates were used to inoculate 5 mls of LB media, containing suitable antibiotics, and grown overnight at 30°C, 200 rpm. This overnight culture was used to seed a 50 mls culture of LB containing suitable antibiotics in a 250 ml flask to an optical density of 0.05 at 600 nm (OD₆₀₀). This culture was grown at 30°C, 200 rpm until the OD₆₀₀ reached 0.4 at which point protein production was induced by the addition of 1 mM IPTG. The cells were then grown for 4 hours at 30°C, 200 rpm and the final OD₆₀₀ measured. The cells were collected by centrifugation and resuspended to an OD₆₀₀ equivalent of 10 in 20 mM sodium phosphate pH 7.4, 20 µg/ml DNase, 0.1 mg/ml egg white lysozyme and frozen. Cells were lysed by freeze-thawing. Where appropriate, the insoluble fraction was removed by centrifugation and the soluble fraction removed quickly to a new container.

E.coli strains which contain pKEHS1208 or pKEHS1209 in addition to the expression vectors for derivatives of tPA were streaked out from glycerol stocks stored at -70°C onto agar plates containing suitable antibiotics to allow for selection. The next day one colony from these plates were used to inoculate 5 mls of LB media, containing suitable antibiotics, and grown overnight at 30°C, 200 rpm. This overnight culture was used to seed a 50 mls culture of LB containing suitable antibiotics in a 250 ml flask to an optical density of 0.05 at 600 nm (OD₆₀₀). This culture was grown at 30°C, 200 rpm until the OD₆₀₀ reached 0.4 at which point pre-induction of Erv1p / PDI or Erv1p / PDI was induced by the addition of 0.5 % arabinose. After 30 minutes production of the tPA derivatives was induced by the addition of 1 mM IPTG. The cells were then grown for 4 hours at 30°C, 200 rpm and the final OD₆₀₀ measured. The cells were collected by centrifugation and resuspended to an OD₆₀₀ equivalent of 10 in 20 mM so-

dium phosphate pH 7.4, 20 µg/ml DNase, 0.1 mg/ml egg white lysozyme and frozen. Cells were lysed by freeze-thawing. Where appropriate, the insoluble fraction was removed by centrifugation and the soluble fraction removed quickly to a new container.

- 5 **Figure 10** shows analysis of human kringle tPA in various systems in LB media at 30°C. Panel A) SDS-PAGE from expression of the human kringle tPA as a fusion protein with MBP in the cytoplasm of *E.coli*. BL21 pLysS RARE (BL) or rosetta-gami (RG), with co-expression with a sulfhydryl oxidase (SO) and DsbC. Lanes marked with (T) and (S) show the total *E.coli* lysates and the soluble fractions, respectively.
- 10 The position of the MBP-kringle tPA fusion protein is marked with an arrow. There is a clear correlation between soluble expression of soluble fusion protein and activity. Panels B-F) Activity measurements for the activity of human kringle tPA expressed in the cytoplasm of *E.coli* using the Roche chromogenic substrate for tPA. 10 µl of *E.coli* lysate were used for the activity measurements. The intrinsic protease activity of the
- 15 *E.coli* lysate has been subtracted. The values are cited as relative activity (%) compared with the most active system which is origami + pre-expression of the sulfhydryl oxidase and DsbC and co-expression of the sulfhydryl oxidase and PDI. Panel B) shows the results for N-terminal hexa-histidine tagged kringle t-PA. Panels C-F) show the results for the MBP-kringle t-PA fusion protein. Overall note the increased expression of active tPA with i) co-expression of a disulfide isomerase; ii) use of the rosetta-
- 20 gami background; iii) use of a MBP-Kringle tPA fusion protein. Strains used were BL = BL21 (DE3) pLysSRARE; RG = rosetta gami; K2 = Origami with pre-expression of the sulfhydryl oxidase and PDI for one hour before induction of kringle t-PA; K3 = Origami with pre-expression of the sulfhydryl oxidase and DsbC for one hour before induction of kringle t-PA. SO indicates co-expression of the sulfhydryl oxidase, +D indicates co-expression of DsbC, +P indicates co-expression of PDI.

- Expression of tPA or kringle tPA (similar but not identical to Retavase) in the cytoplasm of *E.coli* resulted in the production of insoluble inclusion bodies. Co-expression of a sulfhydryl oxidase increased the activity of kringle tPA significantly (see **figure**
- 30 **10**). Combining this with co-expression of an isomerase, PDI or DsbC and/or with expressing kringle tPA as a fusion protein with Maltose binding protein (MBP) or using the rosetta-gami or origami strains further increased the level of active protein produced (see **figure 10**) and the amount of soluble protein produced (see **figure 10**).

In combination our invention increases production of active protein over 100-fold in a wild-type *E.coli* background and gives circa 3 fold higher tPA activity than the best combination in the commercial rosetta-gami strain without expression of the sulfhydryl oxidase.

5 **Example 7**

Efficient production of soluble folded BPTI and disulfide mutants therein

Bovine pancreatic trypsin inhibitor (BPTI) is a low molecular weight inhibitor of trypsin, kallikrein, chymotrypsin and plasmin. It is available under the name Trasylol and is used to inhibit coagulation during bypass surgery. It is a small protein which contains
10 three non-sequential disulfide bonds. The folding intermediates of BPTI are of academic interest to understand protein folding pathways and especially the roles of protein folding catalysts in these processes.

pHEE12, is a derivative of pET23a which has cloned into the multi-cloning site a gene encoding for an N-terminal hexa-histidine tag (MHHHHHHM) followed in frame with
15 the mature form of Bovine pancreatic trypsin inhibitor (BPTI) as represented by the fragment Arg 36 - Ala 93 of the full length protein. This vector expresses BPTI upon induction with IPTG.

pHEE8, is a derivative of pHEE12 in which Cysteine 49 and Cysteine 73 have been mutated to Alanine. This vector expresses C49A/C73A BPTI upon induction with
20 IPTG.

pHEE10, is a derivative of pHEE12 in which Cysteine 65 and Cysteine 86 have been mutated to Alanine. This vector expresses C65A/C86A BPTI upon induction with IPTG.

pKEHS1208 is a derivative of pLysS in which the gene for sulfhydryl oxidase Erv1p
25 from *S.cerevisiae* (Met 1 – Glu 189) (SEQ ID NO: 5) and the gene for the mature form of PDI from *H.sapiens* (Asp 18 – Leu 508) (SEQ ID NO: 11) have been cloned under an arabinose inducible promoter / terminator system. This vector allows pre-expression of Erv1p and PDI upon induction with arabinose.

pKEHS1209 is a derivative of pLysS in which the gene for sulfhydryl oxidase Erv1p
30 from *S.cerevisiae* (Met 1 – Glu 189) (SEQ ID NO: 5) and the gene for the mature form

of DsbC from *E.coli* (Asp 21 – Lys 236) (SEQ ID NO: 10) have been cloned under an arabinose inducible promoter / terminator system. This vector allows pre-expression of Erv1p and DsbC upon induction with arabinose.

5 *E.coli* strains with these expression vectors (excluding strains which contain pKEHS1208 or pKEHS1209) were streaked out from glycerol stocks stored at -70°C onto agar plates containing suitable antibiotics to allow for selection. The next day one colony from these plates were used to inoculate 5 mls of LB media, containing suitable antibiotics, and grown overnight at 30 °C, 200 rpm. This overnight culture was used to seed a 50 mls culture of LB containing suitable antibiotics in a 250 ml flask to
10 an optical density of 0.05 at 600 nm (OD₆₀₀). This culture was grown at 30 °C, 200 rpm until the OD₆₀₀ reached 0.4 at which point protein production was induced by the addition of 1 mM IPTG. The cells were then grown for 4 hours at 30 °C, 200 rpm and the final OD₆₀₀ measured. The cells were collected by centrifugation and resuspended to an OD₆₀₀ equivalent of 10 in 20 mM sodium phosphate pH 7.4, 20 µg/ml DNase, 0.1
15 mg/ml egg white lysozyme and frozen. Cells were lysed by freeze-thawing. Where appropriate, the insoluble fraction was removed by centrifugation and the soluble fraction removed quickly to a new container.

E.coli strains which contain pKEHS1208 or pKEHS1209 in addition to the expression vectors for derivatives of BPTI were streaked out from glycerol stocks stored at -70°C
20 onto agar plates containing suitable antibiotics to allow for selection. The next day one colony from these plates were used to inoculate 5 mls of LB media, containing suitable antibiotics, and grown overnight at 30 °C, 200 rpm. This overnight culture was used to seed a 50 mls culture of LB containing suitable antibiotics in a 250 ml flask to an optical density of 0.05 at 600 nm (OD₆₀₀). This culture was grown at 30 °C, 200 rpm
25 until the OD₆₀₀ reached 0.4 at which point pre-induction of Erv1p / PDI or Erv1p / DsbC was induced by the addition of 0.5 % arabinose. After 30 minutes production of the BPTI derivatives was induced by the addition of 1 mM IPTG. The cells were then grown for 4 hours at 30 °C, 200 rpm and the final OD₆₀₀ measured. The cells were collected by centrifugation and resuspended to an OD₆₀₀ equivalent of 10 in 20 mM so-
30 dium phosphate pH 7.4, 20 µg/ml DNase, 0.1 mg/ml egg white lysozyme and frozen. Cells were lysed by freeze-thawing. Where appropriate, the insoluble fraction was removed by centrifugation and the soluble fraction removed quickly to a new container.

When N-terminally hexa-histidine tagged wild-type mature BPTI or the C49A/C73A or C65A/C86A mutants, which mimic two of the two disulfide bond containing intermediates, are expressed in the cytoplasm of *E.coli* they form inclusion bodies. In contrast, when they are co-expressed with a sulfhydryl oxidase and a thiol-disulfide isomerase soluble protein is formed (see **figure 11**). Purification of the C49A/C73A or C65A/C86A mutants using immobilized metal affinity chromatography and subsequent analysis by reverse phase HPLC indicated that the purified proteins were homogenous, with the elution point being equivalent to the relevant two disulfide containing wild-type folding intermediate (see **figure 11**) i.e. only a single disulfide bond containing species which contained the correct disulfides was obtained.

Claims

1. A method for producing a protein of interest containing one or more disulfide bonds in its native state, which comprises that a prokaryotic host cell is genetically engineered to express the protein of interest and a sulfhydryl oxidase in the cytoplasm of the host cell, said protein of interest being formed in a soluble form and containing disulfide bonds due to the presence of the sulfhydryl oxidase in the cytoplasm of said host cell.
2. The method according to claim 1, wherein the protein product comprising the protein of interest is recovered from the cell culture or from the host cells and optionally purified.
3. The method according to claim 1 or 2, wherein the protein of interest is co-expressed with the sulfhydryl oxidase.
4. The method according to any one of claims 1 to 3, wherein the sulfhydryl oxidase is expressed prior to the protein of interest.
5. The method according to any one of the preceding claims, wherein the host cells are engineered to further express a thiol-disulfide isomerase in the cytoplasm.
6. A prokaryotic host cell for producing a protein of interest containing natively folded disulfide bonds, which comprises that the host cell is genetically engineered to express a sulfhydryl oxidase and a protein of interest in the cytoplasm of the host cell.
7. The method or the host cell according to any one of the preceding claims, wherein the sulfhydryl oxidase uses FAD as a cofactor and contains one, or more, redox active disulfide bonds.
8. The method or the host cell according to any one of the preceding claims, wherein the sulfhydryl oxidase belongs to sulfhydryl oxidases comprising ERV/ALR sulfhydryl oxidase domain.
9. The method or the host cell according to any one of the preceding claims, wherein the sulfhydryl oxidase belongs to ERO family of sulfhydryl oxidases.

10. The method or the host cell according to any one of the preceding claims, wherein the host cell further expresses a thiol-disulfide isomerase in the cytoplasm.
11. The method or the host cell according to claim 10, wherein the disulfide isomerase is PDI, or DsbC.
- 5 12. The method or the host cell according to any one of the preceding claims, wherein the prokaryotic host is a gram negative bacterium, such as *E. coli*.
13. The method or the host cell according to any one of the preceding claims, wherein the protein of interest is produced as a fusion protein.
- 10 14. The method or the host cell according to any one of the preceding claims, wherein the sulfhydryl oxidase and the disulfide isomerase are induced either separately or together with the protein of interest.
15. The method or the host cell according to any one of the preceding claims, wherein the host cell is deficient in thioredoxin reductase and/or glutathione reductase activity.
- 15 16. A vector system for a prokaryotic host cell for producing a protein of interest containing natively folded disulfide bonds, which comprises
- a vector encoding a protein of interest, or having a site for a nucleic acid sequence encoding a protein of interest, and a sulfhydryl oxidase,
- or
- 20 - a first vector encoding a protein of interest, or having a site for a nucleic acid sequence encoding a protein of interest, and
- a second vector encoding a sulfhydryl oxidase,
- wherein said vector or vectors are constructed to be capable of expressing the protein of interest and the sulfhydryl oxidase in the cytoplasm of a prokaryotic host cell.
- 25 17. The vector system according to claim 16 for producing a protein of interest, wherein the system further comprises a vector encoding a thiol-disulfide isomerase, said vector being the same or different vector as the vectors in the vector system, and

said vector being constructed to be capable of expressing a thiol-disulfide isomerase in the cytoplasm of the prokaryotic host.

18. The vector system according to claim 16 or 17, wherein the sulfhydryl oxidase and thiol-disulfide isomerase are induced either separately or together with the protein of interest.

19. A prokaryotic host cell , which comprises the vector system according to any one of claims 16 to 18.

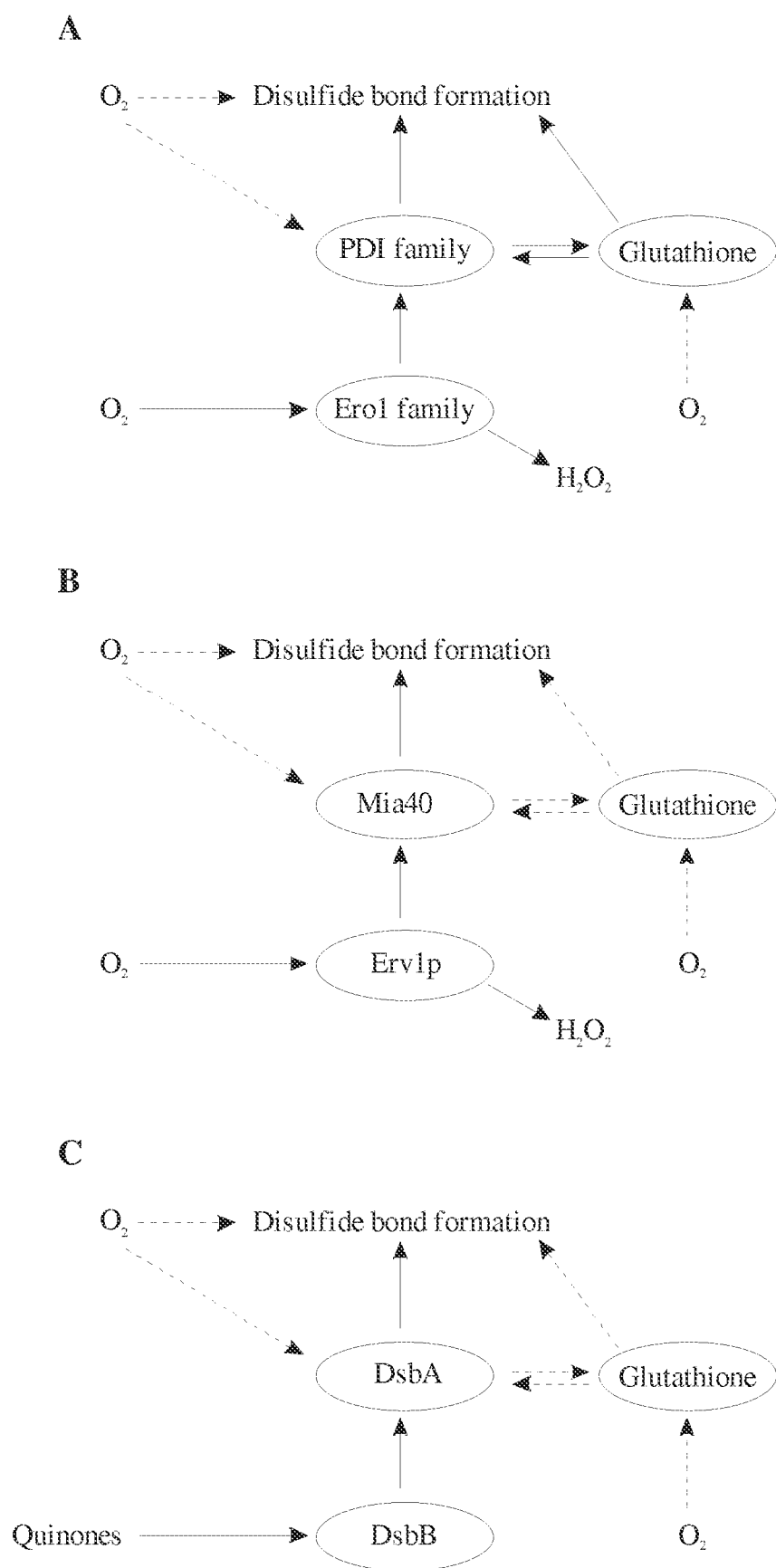


Fig. 1

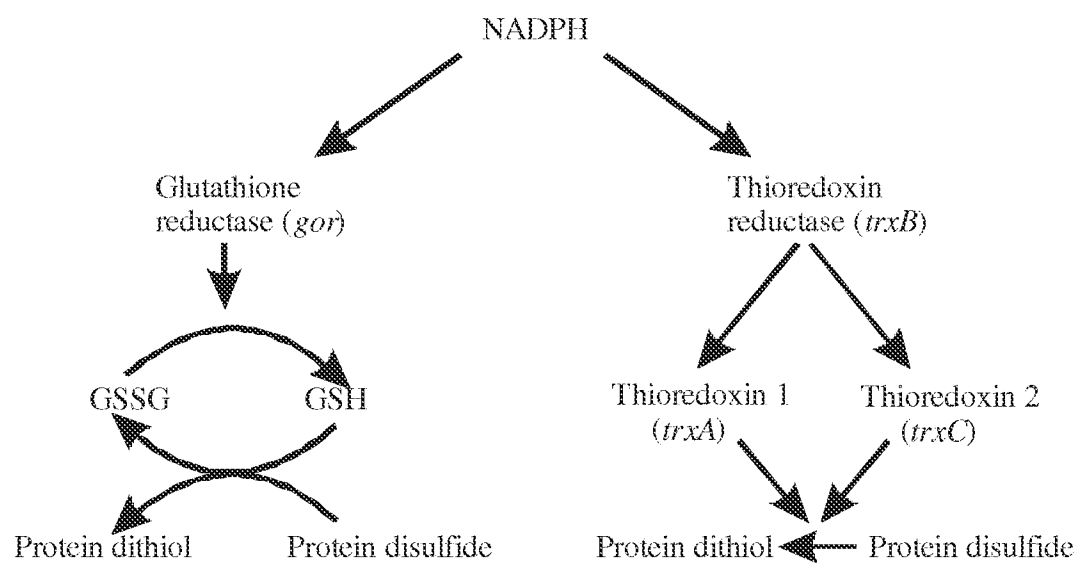


Fig. 2

Endoplasmic oxidoreductin-1 from *Saccharomyces cerevisiae* (Swiss-Prot ID Q03103) (SEQ ID NO:1):

MRLRTAIATLCLTAFTSATSNNSYIATDQTQNAFNDTHFCKVDRNDHVSPSCNVTFNELNAINENIRDDLSALLKSDFFK
YFRLDLYKQCSFWDANDGLCLNRACSV DVVEDWDTLPEYWQPEILGSFNNDTMKEADDSDECKFLDQLCQTSKKPVDIE
DTINYCDVNDFNGKNAVLIDLTANPERFTGYGGKQAGQIWSTIYQDNCFTIGETGESLAKDAFYRLVSGFHASIGTHLSK
EYLNTKTGKWEPNLDFMARIGNFPDRVTNMYFNYAVVAKALWKIQPYLPEFSFCDLVNKEIKNKMDNVISQLDTKIFNE
DLVFANDLSLTLKDEFRRFKNVTKIMDCVQCDCRLWGKIQTGTGYATALKILFEINDADEFTKQHIVGKLTKEYELIALL
QTFGRLESIESVNMFEKMYGKRLNGSENRLSSFFQNNFFNILKEAGKSIRYTIENINSTKEGKKKTNNSSQSHVFDDLKM
PKAEIVPRPSNGTVNKWKKAWNTEVNNVLEAFRFIYRSYLDLPRNIWELSLMKVYKFWNKFIGVADYVSEETREPI SYKL
DIQ

Human ERO1-like protein alpha (Swiss-Prot ID Q96HE7) (SEQ ID NO:2):

MGRGWGFLFGLLGAVWLLSSGHGEEQPPETAAQRCFCQVSGYLDDCTCDVETIDRFNNYRLFPRLQKLLESDFRYKYVN
LKRPCPFWNDISQCGRRDCAVKPCQSDEVPDGIKSASYKYSEEANNLIEECEQAERLGAVDESLSEETQKAVLQWTKHDD
SSDNFCEADDIQSPEAEYVDLLNPERYTGKGPDAWKIWNVIYEENCFKPQTIKRPLNPLASGQGTSEENTFYSWLEGL
CVEKRAFYRLISGLHASINVHLSARYLLQETWLEKKWGHNITEFQQRFDGILTEGEGPRRLKNLYFLYLIELRALSKVLP
FFERPDLFTGNKIQDEENKMLLLEILHEIKSFPLHFDENSFFAGDKKEAHKLKEDFRLHFRNISRIMDCVGCFCRLW
GKLQTQGLGTALKILFSEKLIANMPESGPSYEFHLTRQEIVSLFNAFGRISTSVKELENFRNLLQNIH

Human ERO1-like protein beta (Swiss-Prot ID Q86YB8) (SEQ ID NO:3):

MSQGVRRAGAGQGVAAAVQLLVTL SFLRSVVEAQVTGVLDDCLCDIDSIDNFNTYKIFPKIKKLQERDYFRYYKVNLRP
CPFWAEDGHCSIKDCHVEPCPESKIPVGIKAGHSNKYLKMANNTKELEDCEQANKLGAINSTLSNQSKEAFIDWARYDDS
RDHFCELDDERSPAAQYVDLLNPERYTGKGTSAWRVWNSIYEENCFKPRSVYRPLNPLAPSRGEDDGESFYTWLEGLC
LEKRVFYKLISGLHASINLHLCANYLLEETWGKPSWGPNIKEFKHRFDPVETKGEPRRLKNLYFLYLIELRALSKVAPY
FERSIVDLYTGNAEEDADTKTLLNIFQDTKSFPMHFDEKSMFAGDKKGAKSLKEEFRLHFKNISRIMDCVGCDCRLWG
KLQTQGLGTALKILFSEKEIQKLPENSPSGFQLTRQEIVALLNAFGLSTSI RDLQNFVLLQHSR

Fig. 3

Human ALR (Swiss-Prot ID P55789) (SEQ ID NO:4):

MAAPGERGRFHGGNLFFLPGGARSEMDDLATDARGRGAGRRDAAASASTPAQAPTSDSPVAEDASRRRPCRACVDFKTW
MRTQQKRDTKFREDCPPDREELGRHSWAVLHTLAAYYPDLPTPEQQQDMAQFIHLFSKFYPCEECAEDLRKRLCRNHDPDT
RTRACFTQWLCHLHNEVNRKLGKPDFDCSKVDERWRDGGWKGSCD

S. cerevisiae mitochondrial FAD-linked sulfhydryl oxidase ERV1 (Swiss-Prot ID P27882) (SEQ ID NO:5):

MKAIDKMTDNPPQEGLSGRKIIYDEDGKPCRSCNTLLDFQYVTGKISNGLKNLSSNGKLAGTGALTGEASELMPGSRTYR
KVDPPDVEQLGRSSWTLHLSVAASYPAOPTDQOKGEMKQFLNIFSHIYPCNWCAKDFEKYIRENAPQVESREELGRWMCE
AHNKVNKKLRKPKFDCNFWEKRWKDGWDE

S. cerevisiae FAD-linked sulfhydryl oxidase ERV2 (Swiss-Prot ID Q12284) (SEQ ID NO:6).

MKQIVKRSHAIRIVAALGIIGLWMFFSSNELSIATPGLIKAKSGIDEVQGAAAEKNDARLKEIEKQTIMPLMGDDKVKKE
VGRASWKYFHTLLARFPDEPTPEEREKLHTFIGLYAELYPGCECSYHFVKLIEKYPVQTSRTAAAMWGCHIHNKVNEYL
KKDIYDCATILEDYDCGCSDSDGKRVSLKEAKQHG

Human sulfhydryl oxidase 1 (QSOX; Swiss-Prot ID O00391) (SEQ ID NO:7):

MRRCNSSGSGPPPSLLLLLLWLLAVPGANAAPRSALYSPSDPLTLLQADTVRGAVLGSRSAAWAVEFFASWCGHCIAFAPTW
KALAEDVKAWRPALYLAALDCAETNSAVCRDFNIPGFPTVRFFKAFTKNGSGAVFPVAGADVQTLRERLIDALESHHDT
WPPACPPLEPAKLEEIDGFFARNNEEYLALIFEKGGSYLGREVALDLSQHKGVAVRRVLNTEANVVRKFGVTDFPSCYLL
FRNGSVSRVPVLMESRSFYTAYLQRLSGLTREAAQTTVAPTTANKIAPTVMWKLADRSKIYMADLESALHYILRIEVGRFP
VLEGQRLVALKKFVAVLAKYFPGRPVQNLHNSVNEWLKRQKRNI PYSFFKTALDDRKEGAVLAKKVNWIGCQGSEPHF
RGFP CSLWVLFHFLT VQAARQNV DHSQEAAKAKEVLP AIRGYVHYFFGCRDCASHFEQMAAASMHVRV GSPNA AVLWLWSS
HNRVNARLAGAPSED P QFPKVQWPPRELCSACHNERLDVPVWDVEATL N FLKAHFSPSNI ILDFPAAGSAARRDVQNVAA
APELAMGALELESRNSTLDPGKPEMMKSPTNTTPHVPAEGPEASRPPKLHPGLRAAPGQEPPEHMAELQRNEQEQLGQW
HLSKRD TGAALLAESRAEKNRLWGP LEVRRVGRSSKQLVDIPEGQLEARAGRGRGWLQVLGGGFSYLDISLCVGLYSL S
FMGLLAMYTYFQAKIRALKGHAGHPAA

Fig. 4A

Human sulfhydryl oxidase 2 (Swiss-Prot ID Q6ZRP7) (SEQ ID NO:8):

MAAAGAAVARSPGIGAGPALRARRSPPPRARLPRLLVLLAAAVGPGAGGAARLYRAGEDAVWVLDSGSGVRGATANSSA
AWLVQFYSSWCGHCIGYAPTWRALAGDVRDWSAIRVAALDCMEEKNQAVCHDYDIHFYPTFRYFKAFTKEFTTGENFKG
PDRELRTVRQTMIDFLQNHTEGSRPPACPRLDPIQPSDVLSLLDNRGSHYVAIVFESNSSYLGREVILDLIPYESIVVTR
ALDGDKAFLEKLGVSVPSCYLIYPNGSHGLINVVKPLRAFFSSYLKSLPDVRKKSLLPEKPHKEENSEIVVWREFDKS
KLYTVDLESGLHYLLRVELAAHKSLAGAELKTLKDFVTVLAKLFPGRPPVKKLLEMLQEWLASLPLDRIPYNAVLDLVNN
KMRISGIFLTNHIKWVGCQGSRSSELRGYPCSLWKLFHTLTVEASTHPDALVGTGFEDDPQAVLQTMRRYVHTFFGCKECG
EHFEEMAKESMDSVKTPDQAILWLWKKHNMVNGRLAGHLSLSEDPRFPKLQWPTPDLCPACHEEIKGLASWDEGHVLTFLKQ
HYGRDNLLDTYSADQGDSSSEGGTLARGEETKRLTPPEVSHGDRDTQSVRPPGALGPRPALPESLHHSLDGKLQSLDGPG
AHKEVGGAAPFLGVDFSSLDMSLCVVLVYVASSFLVMVYFFFRVRSRRWKVKKHHHPAV

Vaccinia virus FAD-linked sulfhydryl oxidase E10 (Swiss-Prot ID P21050) (SEQ ID NO:9)

MNPKHWGRAVWTIIIFIVLSQAGLDGNI EACKRKLYTIVSTLPCPACRRHATIAIEDNNVMSSDDLNYIYYFFIRLFNNLA
SDPKY AIDVSKVKPL

E.coli (K12 strain) DsbC (SEQ ID NO:10) :

MKKGFMLFTLLAAFSGFAQADDAAIQQTLAKMGIKSSDIQPAPVAGMKTVLTN SGVLYITDDGKHIIQGPMYDVSGTAP
VNVTNKMLLKQLNALEKEMIVYKAPQEKHVITVFTDITCGYCHKLHEQMADYNALGITVRYLAFPRQGLDSDAEKEMK
AIWCAKDKNKAFFDDVMAGKSVAPASCDVDIADHYALGVQLGVSGTPAVVLSNGTLVPGYQPPKEMKEFLDEHQKMT
SGK

Human PDI (SEQ ID NO:11):

MLRRALLCLAVAALVRADAPEEEEHVLVLRKSNFAEALAAHKYLLVEFYAPWCGHCKALAPEYAKAAGKLKAEGSEIR
LAKVDATEESDLAQYGVGYPTIKFFRNGDTASPKEYTAGREADDIVNLKKRTGPAATTL PDGAAAESLVESSEV
AVIGFFKDVEDSAKQFLQAAEAIDDIPFGITSNSDVFSKYQLDKDGVVLFKKFDEGRN NFEGEVTKENLLDFIKHNQL
PLVIEFTEQTAPKIFGGEIKTHILLFLPKSVSDYDGKLSNFKTAAESFGKILFIFIDSDHTDNQRILEFFGLKKEECPAVRL
ITLEEEMTKYKPESEELTAERITEFCHRFLEGKIKPHLMSQELPEDWDKQPVKVLVGKNFEDVAFDEKKNVFVEFYAP
WCGHCKQLAPIWDKLGETYKDHENIVIAKMDSTANEVEAVKVHSFPTLKFFPASADRTVIDYNGERTLDGFKKFLESG
GQDGAGDDDDLEDLEEAEEPDMEEDDDDQKAVKDEL

Fig. 4B

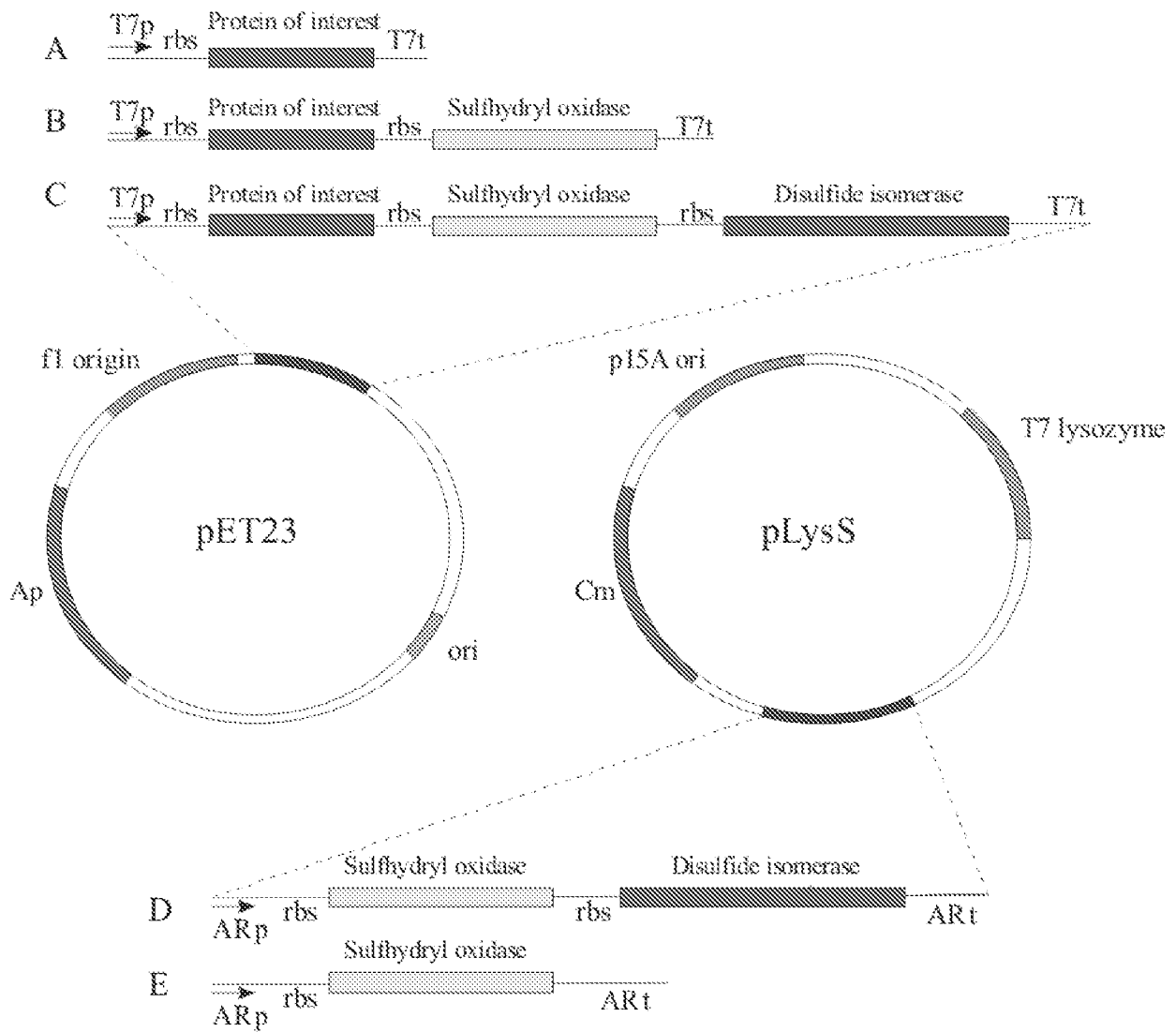


Fig. 5

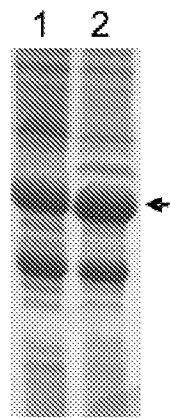


Fig. 6

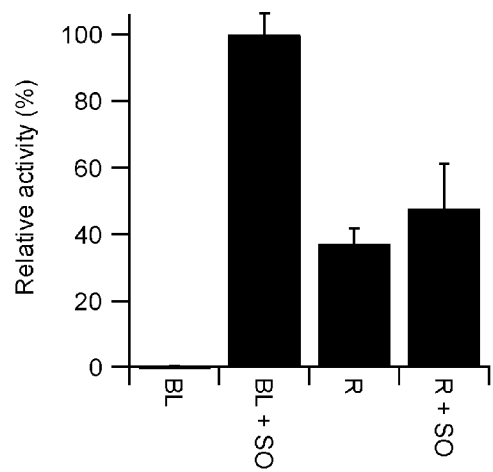


Fig. 7

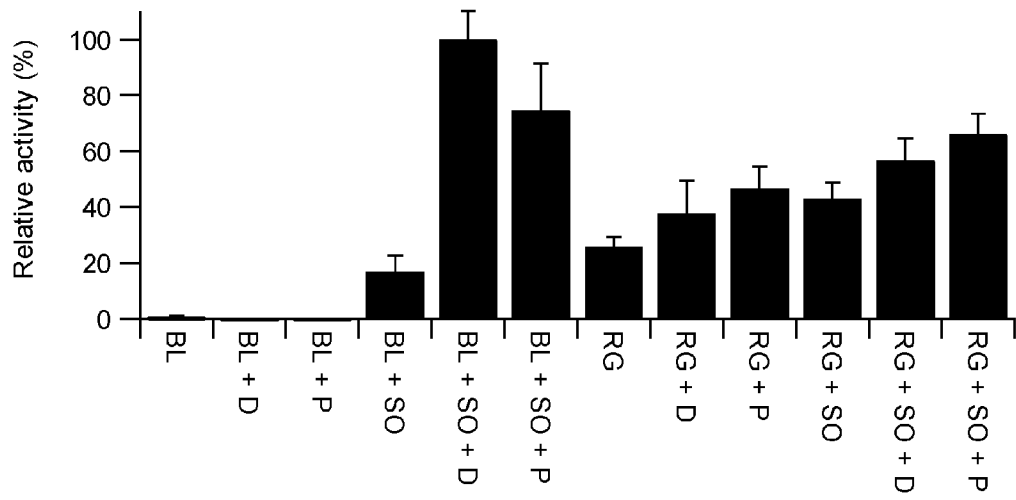


Fig. 8

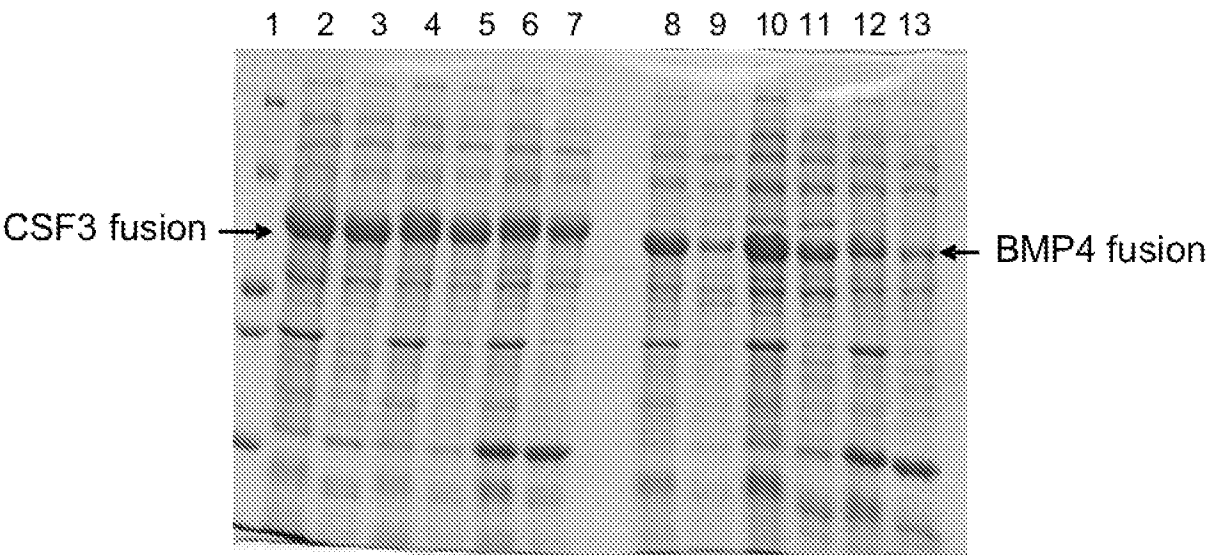


Fig. 9

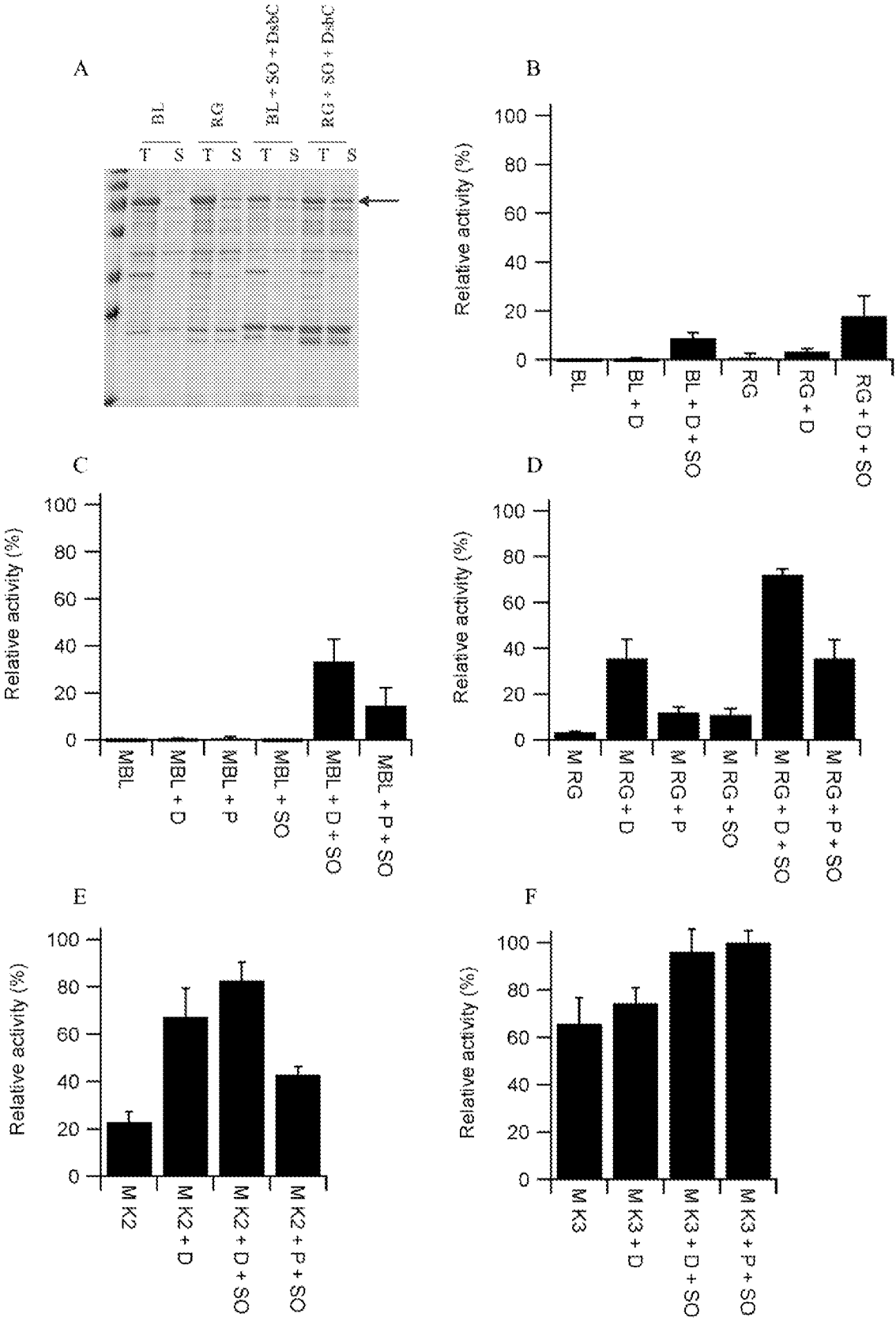


Fig. 10

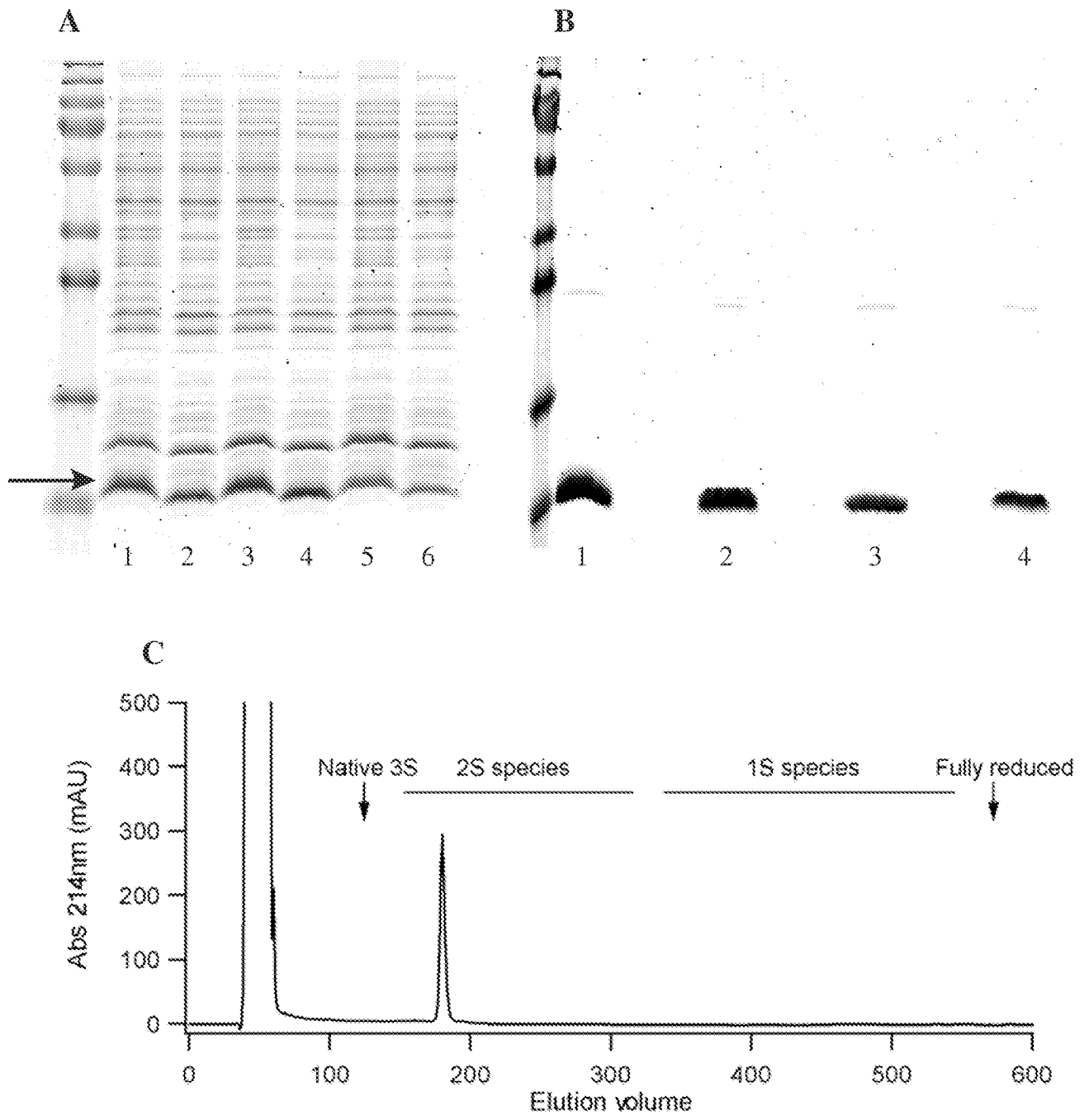


Fig. 11

INTERNATIONAL SEARCH REPORT

International application No

PCT/FI2010/050448

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12N15/70 C12N9/02
 ADD. C12N15/00 C12N9/00 C07K14/00 C12N5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data, CHEM ABS Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99/07727 A1 (MASSACHUSETTS INST TECHNOLOGY [US]) 18 February 1999 (1999-02-18) page 1 - pages 5, 8; claims 24-26 -----	1-19
X	WO 2009/058956 A1 (DANISCO US INC [US]; BAO KAI [US]; WANG HUAMING [US]) 7 May 2009 (2009-05-07) claim 1; figures 1-5 -----	1-19
A	WO 2005/061718 A1 (DELTA BIOTECHNOLOGY LTD [GB]; SLEEP DARRELL [GB]; SHUTTLEWORTH GILLIAN) 7 July 2005 (2005-07-07) the whole document ----- -/--	1-19



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

4 August 2010

Date of mailing of the international search report

13/08/2010

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040.
 Fax: (+31-70) 340-3016

Authorized officer

Seranski, Peter

INTERNATIONAL SEARCH REPORT

International application No

PCT/FI2010/050448

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>VITU E ET AL: "Gain of Function in an ERV/ALR Sulfhydryl Oxidase by Molecular Engineering of the Shuttle Disulfide" JOURNAL OF MOLECULAR BIOLOGY, LONDON, GB LNKD- DOI:10.1016/J.JMB.2006.06.070, vol. 362, no. 1, 8 September 2006 (2006-09-08), pages 89-101, XP024951351 ISSN: 0022-2836 [retrieved on 2006-09-08] the whole document -----</p>	1-19

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/FI2010/050448

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9907727	A1	18-02-1999	NONE
WO 2009058956	A1	07-05-2009	AU 2008318644 A1 07-05-2009 CA 2704548 A1 07-05-2009 EP 2203556 A1 07-07-2010 US 2009221030 A1 03-09-2009
WO 2005061718	A1	07-07-2005	AT 421994 T 15-02-2009 AU 2004303601 A1 07-07-2005 AU 2008202538 A1 26-06-2008 CA 2551503 A1 07-07-2005 CN 1922325 A 28-02-2007 EP 1709181 A1 11-10-2006 EP 2048236 A1 15-04-2009 ES 2321948 T3 15-06-2009 GB 2424889 A 11-10-2006 HK 1099938 A1 24-12-2009 JP 2007515962 T 21-06-2007 KR 20070005568 A 10-01-2007 SG 149003 A1 29-01-2009 US 2007275889 A1 29-11-2007 ZA 200605588 A 25-02-2009